DOCUMENT RESUME

ED 069 702 24 TM 002 154

AUTHOR Emrick, John A.
TITLE The Experimental Validation of an Evaluation Model for Mastery Testing. Final Report.
INSTITUTION Massachusetts Univ., Amherst.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
BUREAU NO BR-0-A-063
PUB DATE Nov 71
GRANT OEG-1-71-0002
NOTE 100p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Achievement Tests; *Cognitive Measurement; Item Analysis; *Mathematical Models; Measurement Techniques; Research Methodology; *Test Construction; *Test Validity

ABSTRACT

The validity of an evaluation model for mastery testing applications was investigated. Three variables were tested in an experiment using 96 third grade subjects--amount of training, number of alternates in an item, and number of items. The concept hierarchy involved an orderly progression from a concept involving one relevant of three varying dimensions through two relevant of four varying dimensions (concept 2) to four relevant of six varying dimensions (concept 3). This established the basis for computing mastery evaluation cut rules on the basis of the model. Reliable differences occurred for training level and for concept difficulty, but not for test length or item types. The results of the validity analysis were, in general, favorable to the model. It is thus concluded that the proposed model is reasonably valid. This evidence could be used as a basis for a demonstration or experimental implementation of the model in an educational environment that uses mastery evaluation procedures. (DJ)

FINAL REPORT
Project No. O-A-063
Grant No. OEG-1-71-0002

THE EXPERIMENTAL VALIDATION OF AN EVALUATION MODEL FOR MASTERY TESTING

John A. Emrick
University of Massachusetts
Amherst, Massachusetts 01002

November 1971

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Final Report
Project No. 0-A-063
Grant No. OEG-1-71-0002

THE EXPERIMENTAL VALIDATION OF AN EVALUATION MODEL FOR MASTERY TESTING

John A. Emrick
Stanford Research Institute
Menlo Park, California 94025

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

ACKNOWLEDGMENTS

The author wishes to express his appreciation and gratitude to all of those who assisted in the completion of this research. Particular thanks are extended to Mr. Sergio de Luna and Ms. Patricia Dede who diligently assisted in the preparation and execution of the study.

The close cooperation and assistance provided by the administrators and teachers of the schools involved deserves special recognition. The author is especially indebted to the following individuals: Mr. Warren D. Bennet, Superintendent of Schools, Frontier Regional School District; Mr. Paul Aronstam, principal, and Ms. Pat Scott, teacher, East Whately Elementary School; Mr. Norm Halliwell, coordinator of elementary education, Greenfield School District; Mr. Hugh Hayden, principal, and Ms. Brown, Hinrich, Osgood, Shea, and Melaven, teachers, Four Corners Elementary School; Mr. Chris Grammaticus, Superintendent of Schools, and Mr. Ralph Meecham, elementary supervisor, West Springfield School District; Ms. Genevieve Cocchi, principal, and Ms. Kaczman and De Feo, teachers Park Avenue Elementary School; Mr. Paul Dion, principal, and Ms. Donna Corley, teacher, Fousely Elementary School.

Finally, special thanks are given to Ms. Diane Hogland for her valuable support and assistance provided in the preparation of materials and processing of data, and to Ms. Trudy Hammerling for her skilled assistance in the production and revision of this research report.

SUMMARY

The research described in this report was designed to investigate experimentally the validity of an evaluation model for mastery testing applications. This model is based on the assumptions that: (1) the learning of fundamental skills can be considered all or none; (2) each item response on a single skill test represents an unbiased sample of the examinee's true mastery status; (3) measurement error occurring on the test (as estimated from the average interitem correlation) can be of only one type (α or β) for each examinee; and (4) through practical and theoretical considerations of evaluation error costs and item error characteristics, an optimal mastery criterion can be calculated. Each of these assumptions is discussed, and the resultant mastery criteria algorithm is presented in a form amenable to experimental validation.

Operationalization of the key parameters of the model was accomplished through use of a train-test design in a concept attainment experiment. Two alternative models of hierarchical concept attainment were used as the basis of materials development and sequencing. The mastery probability parameter was manipulated by varying the extent of training Ss reviewed at each step in the attainment hierarchy. The item error parameter was manipulated by varying the number of alternatives (choices) in the test item. The length parameter was manipulated by using tests of varying length.

The experiment was conducted on 96 third grade Ss. Each S received familiarizations training and then proceeded through the train-test sequences of the experiment. A given S was assigned to either of two train-transfer paradigms (intradimensional shift versus no-shift) under one of three levels of training (low, moderate, high). The moderate training level was the theoretical average number of trials required to attain the concept. The concept hierarchy involved an orderly progression from a concept involving one relevant of three varying dimensions through two relevant of four varying dimensions (concept 2) to four relevant of six varying dimensions (concept 3).

Immediately following training, Ss received concept attainment tests in the form of blank trials. A given S was assigned to either continuous versus terminal testing, and was given either five or 10 item tests that were either two or four choice.

Three groups of analyses were performed: training, testing, and model validation. The results of the analysis of training data validated the assumptions underlying the levels manipulation, and provided support for a cumulative component learning and transfer paradigm (the no-shift sequence). Performance also was shown to interact with the level of training in terms of initial trial solution biases. When this bias was controlled for, the hypothesized outcomes become more apparent.

Analysis of test data failed to reveal any systematic performance differences other than those observed as a function of training manipulations. That is, reliable differences occurred for training level and for concept difficulty, but not for test length or item type. This concurrence of training and test results established the basis for implementation of the mastery evaluation algorithm and for subsequent assessment of its operational validity.

The results of this validity analysis were, in general, favorable to the model. For example, the assumption that the cut rules are optimal, for given measurement and decision error constraints, was supported in that the optimization ratio was computed to be 0.95. It also is shown that the overall test distributions for five and 10 item tests separate into the respective mastery and nonmastery components quite differently on individual than on aggregated bases. Furthermore, the theoretical and empirical distributions of mastery and nonmastery scores show reasonably good fit for the five item tests, although not so good for the longer tests.

It is thus concluded that the experimental evidence provided by this research is reasonably supportive of the validity of the proposed mastery evaluation model. This evidence could be used as a basis for a demonstration or experimental implementation of the model in an educational environment that uses mastery, or criterion referenced, evaluation procedures.

CONTENTS

ILLUSTRATIONS

ILLUSTRATIONS

TABLES

Chapter I

INTRODUCTION

Among the more exciting and promising trends currently emerging with educational innovations and reforms is a shift from traditional classroom instruction with its norm-referenced testing procedures to more individualized instructional systems based on criterion referenced test procedures (e.g., Block, 1971). One particularly successful system is known as Individually Prescribed Instruction (IPI). It places emphasis on materials and methods of instruction, matching these with level of achievement, past progress, and perhaps learning style of the student in generating a work assignment or "prescription" (Cooley and Glaser, 1969; Glaser, 1967). However, as with other individualized systems, IPI is critically dependent on the existence of a reliable and valid measurement model to indicate when the student has attained each skill-mastery state.

It is not difficult to show that the traditional measurement procedures are inadequate, or at best arbitrary as a method of identifying student skill mastery. For example, using criterion referenced procedures, IPI has suggested an 85 percent correct minimum as a mastery criterion for any skill test (of which there are more than 400). Although this criterion does have intuitive appeal, there is no convenient analytical or empirical justification for it. Just as various skills may differ in level of difficulty in terms of mastery, so also might the optimal performance criteria in the test situation vary. It may be that for some skills, a test score of 60 percent is indicative of mastery, whereas for others a score of 90 percent or higher would be required. In brief, the issue is not whether a criterion referenced testing procedure is or is not appropriate to IPI, but rather how and at what level each criterion should be set.

To anchor the skill-testing procedure to the operations and outcomes of individualized instructional technology, a skill-mastery test model is proposed (Emrick, 1971) in which both item and student information are combined, yielding probability statements regarding skill-mastery status. This model is particularly attractive in that the assumptions are few and simple, and it provides for empirical determination of the most critical parameters--namely, the item measurement error likelihoods. Furthermore, the generation of a test cut rule or mastery criterion is provided by an algorithm that is based on both the test properties and a cost-benefit analysis of decision errors.

Model

The basic working assumptions of this mastery test model are as follows:

(1) Appropriately defined educational objectives will consist of, or can generally be analyzed into, collections of unitary and explicitly defined (in terms of performance criteria) skills (Gagné, 1962). For each of these skills, mastery will be a binary (all or none) variable. Thus, for an educational objective to be mastered completely, all component skills must be mastered. Further, the degree or "level" of mastery of the objective will be determined by the proportion or number of these component skills that are mastered.

(2) Tests designed to assess mastery of the component skills within an objective will each consist of collections of test items that are highly homogeneous in terms of content, form, and difficulty level. Thus, within a single skill test, each item response provides an unbiased estimate of the examinee's mastery status with respect to that skill.

(3) Since mastery of each unitary skill is assumed to be an all or none variable, the measurement error for a given examinee on a single such skill will be of only one of two types: (1) type 1 or alpha (α), in which the examinee's responses lead to a mastery conclusion when his true status is nonmastery; or (2) type 2 or beta (β), in which his item responses lead to a nonmastery conclusion when in fact he has mastered the skill in question. Stated differently, an examinee can occupy only one status with respect to the skill being tested--mastery or nonmastery. Test item responses that correspond to the examinee's true status are, by definition, valid (i.e., mastery students "pass" the item and nonmastery students "fail" the item). Test item responses that do not correspond to the examinee's true status are, also by definition, measurement error (i.e., "lucky guesses" and so forth for nonmastery examinees and "careless errors" and so forth for mastery examinees). This situation is represented in Table 1.

Table 1

Item Response Contingencies as a Function of Learning State on a Single Skill Mastery Test

		Observed Response: wrong (w)	Observed Response: correct (c)
True Learning State	mastery (M)	β	$1-\beta$
	nonmastery ($\bar{M}$)	$1-\alpha$	α

α = The probability of a correct response from a nonmastery student (type 1 error)

β = The probability of a wrong response from a mastery student (type 2 error)

$(1-\alpha)$ = The probability of a valid wrong response

$(1-\beta)$ = The probability of a valid correct response

(4) The extent of measurement error in a single skill test can be approximated by calculating the average interitem correlation of examinee responses to the parallel and homogeneous test items. This average interitem correlation provides an unbiased estimate of the reliability of a single item of a unitary skill test. Since reliability is defined as the proportion of total variance that is "true" variance (Lord and Novick, 1968; Ghiselli, 1964), this average interitem correlation can be interpreted as an unbiased estimate of the squared correlation between an examinee's true mastery state and his item response.

This correlation between mastery state and item response can further be interpreted in terms of the two classes of measurement error (α and β) with reference to Table 1. The response contingencies from this four-fold table are calculated in the form of a phi (ϕ) coefficient, indicating the correlation between observed responses and "true" score. One formula for computing ϕ from Table 1 is:

$$\phi = \frac{1 - \alpha - \beta}{\sqrt{1 - (\alpha - \beta)^2}} \quad (1)$$

Since reliability is defined as the squared correlation between fallible and true scores, and since the above expression represents this correlation on the item level, the square of equation (1) is the expression for item reliability on a single skill mastery test.

(5) Because of the presence of at least some measurement error, decision errors correspondingly will accrue regarding determination of examinee status on the skills being measured. A decision-theoretic approach to this problem (Chernof and Moses, 1959) suggests regret resulting from these evaluation errors can be minimized through a cost-benefit analysis of the variables that comprise the evaluative process. Three classes of these variables are:

- Statistical, such as item reliability and test length. As item scores become more reliable, tests of fixed length will yield proportionally fewer evaluative errors. Similarly, for a given item reliability, increasing test length by adding parallel items will operate to differentiate more clearly mastery from nonmastery examinees (Emrick and Adams, 1969). However, it is not completely clear what the costs involved in improving item reliability would be, or, aside from following principles of item construction (Bormuth, 1970), how one actually would manipulate item reliability. Further, increase in overall test reliability is a decreasing function of increase in item reliability or test length (Lord and Novick, 1968).

- The second class of variables is described as curricular. Specific instructional objectives are seen to vary with respect to the importance they occupy in differing instructional models. For objectives viewed as ancillary to the model, evaluative regret will be low or irrelevant. However, for objectives viewed as fundamental to the model and prerequisite to further learning, regret can become sizable.

This regret will accrue from the two types of evaluative (decision) error: (1) Type I--or a "false pass" comprising costs of reduced efficiency in mastery of subsequent objectives due to nonmastery of prerequisite, and if the learner eventually "bogs down," the costs of diagnosis and remediation; and (2) Type II--or a "false fail" comprising costs of unnecessary exercises, materials, and instructional time given this mastery student, as well as the costs of subsequently retesting him.

- The third class of factors that enters into regret is in terms of psychological costs resulting from decision errors. For example, regret for Type I evaluative errors would consist of psychological costs to the "out of track" learner, such as confusion, suboptimal success rate, and the like. The Type II regret would include psychological costs such as boredom, decreased sense of achievement, lower motivation, and so forth.

If meaningful quantitative values could be independently assigned to each of these cost factors, as well as to α and β item error probabilities, then the generation of optimal mastery criteria for a given test would be straightforward. But since no such values are conveniently available (nor are they likely to be in the foreseeable future), Emrick (1971) and Emrick and Adams (1969) have proposed that mastery cutoff scores be optimized in terms of <u>relative</u> decision error costs and <u>relative</u> item error probabilities. Hence the optimization formula:

$$K = \frac{\log \frac{\beta}{1 - \alpha} + 1/n(\log RR)}{\log \frac{\alpha\beta}{(1 - \alpha)(1 - \beta)}} \qquad (2)$$

where

K = the cut point expressed as a percent score on the test

α = estimated probability of Type I item error

β = estimated probability of Type II item error

RR = ratio of regret of Type II to Type I decision errors

n = test length (number of items).

The determination of the values that enter into equation (2) follows from the above described assumption of the model. Specifically, the total probability of item measurement error ($\alpha + \beta$) is estimated as one minus the square root of the average interitem correlation (i.e., $1 - \sqrt{\bar{r}_i}$). A logical analysis of item form yields an estimate of which type of error predominates. For example, a true-false test should yield relatively more α than β errors, whereas for recall or completion items, the reverse should be true. Using these estimates of the sum and ratio of the errors, an estimate of each error component can be obtained.

A similar procedure is used to supply a value for the ratio of regret (RR). Specifically, a logical analysis of the evaluative procedure will yield an estimate of the more costly decision error, possible in conjunction with some estimate of the examinee's status before testing. In some testing situations, false fail errors may be considered far more costly than false passes whereas in other cases the reverse may be true. Also, there may occur cases where the two costs are judged essentially equal. Actually, many teachers operate in this fashion, deciding to err either on the "high" or "low" side, depending on the skill as well as the examinee being evaluated. These estimates are operationalized in equation (2) as RR. Finally, by indicating the test length (n) in addition to the above values, tables of optimal mastery criteria can be generated for virtually any single skills test.

The goals of this research were to validate empirically this evaluation approach to mastery testing. To accomplish this, it is necessary to establish--and to some extent quantify--manipulations among the relevant parameters of the model. These parameters are:

- Mastery probability, or the probability that the student has attained mastery of the skill at the time of testing.
- Item error likelihood, or the relative probabilities of α (false pass) and β (false fail) measurement error occurring in the test.
- The length of the test.

The first of these parameters is the most difficult to directly establish, since it must be derived on the basis of a model of learning, or inferred on the basis of performance. The procedure adopted in this research was to consolidate Gagné's "acquisition of knowledge" hierarchical learning model--versus Neisser and Ween's logical complexity model--with Trabasso and Bower's "discrimination-attention" learning assumptions in the form of type of problems and level of training in concept identification tasks.

Since the second parameter is primarily a function of item type or format and related characteristics (i.e., measurement procedure, which can be brought under experimental control), and since test length is clearly an objective parameter, the major part of the following discussion will deal with the concept identification task.

Concept Identification Task

The subject's task in a concept identification problem is described generally as involving at least two components: the identification of the relevant dimensions, and the identification of the rule or rules that bring the attributes together in a particular fashion (Bruner, Goodnow and Austin, 1956; Haygood and Bourne, 1965; Bourne, 1968).

Given a set of dimensions $\underline{a}$, $\underline{b}$, $\underline{c}$ . . . $\underline{x}$ each with n values or attributes (a1, a2, a3 . . . an; b1, b2, and so forth), the learner's task is to discover or identify the attributes that satisfy the conditions defining the concept. The attributes that satisfy the concept definition are said to be relevant and the dimensions to which they belong are called relevant dimensions. All other attributes that vary, either within or across instances of the concept, are described as irrelevant.

The method and structure by which relevant and irrelevant attributes are arranged determines the conceptual rule. Neisser and Weene (1962) have shown that when the number of relevant attributes is restricted to two, there are 10 such conceptual rules, as summarized in Table 2.

Research on attribute learning has demonstrated the effects of such variables as the number of relevant and irrelevant dimensions (Walker and Bourne, 1961) and the amount of intra- and inter-dimensional variability (Battig and Bourne, 1961). For example, Battig and Bourne's (1961) investigation on the effects on error rate of changes in the number of dimensions revealed that college students made more errors following both inter- and intra-dimensional variations. Further, this relationship between error rate and intradimensional variability was found to correspond closely to a straight line function.

The amount of irrelevant and relevant information also has been shown to contribute to task complexity. Although it would seem on an intuitive basis that increased relevant information should increase the difficulty of the conceptual task, it is not so obvious that increased irrelevant information should do so. Actually, the amount of irrelevant information affects only the complexity of the stimulus pattern, since

Table 2

Conceptual Rules Describing Partitions of a Population with Two Focal Attributes (Red and Square)

Affirmation	All red patterns are examples of the concept.
Conjunction	All red <u>and</u> square patterns are examples.
Inclusive disjunction	All patterns which are red <u>and</u> square <u>or both</u> examples.
Conditional	<u>If</u> a pattern is red <u>then</u> it must be square to be an example.
Biconditional	Red patterns are examples <u>if and only if</u> they are square.
Negation	All patterns which are <u>not</u> red are examples of the concept.
Alternative denial	All patterns which are <u>either not</u> red <u>or not</u> square are examples.
Joint denial	All patterns which are <u>neither</u> red <u>nor</u> square are examples.
Exclusion	All patterns which are red <u>and not</u> square are examples.
Exclusive disjunction	All patterns which are red <u>or</u> square <u>but not</u> both are examples

Modified from Haygood and Bourne, 1965.

the number and type of categories into which the patterns must be sorted will remain the same. Further, Walker and Bourne's (1961) study indicated an interaction between the amount of both relevant and irrelvant information and problem difficulty. Errors increased at a positively accelerated rate with increases in relevant information, but this effect depended on the level of irrelevant information employed in a problem.

Neisser and Weene (1962) demonstrated that rules are not of equal learning difficulty even though they refer to the same set of attributes. Neisser and Weene further showed that the different rules fall logically into three categories or levels based on the number of component elements. Their results indicated that the degree of difficulty of each rule increased from level to level. Although the results do not seem surprising, it is not immediately clear why the different rules would distribute themselves along this continuum of difficulty.

Several explanations have been offered in an attempt to explain why certain rules are more difficult to obtain than others. One possibility suggested by Haygood and Bourne (1965) is that subjects are forming and testing various rule hypotheses until the correct one is discovered. Thus, as concept increases in complexity, more rules become available, reducing the probability of an early solution. This explanation is similar to, if not the same as, the decision tree model suggested by Hunt (1962, cited by Haygood and Bourne, 1965). In a study reported by Neisser and Weene (1962) this assumption of availability of rules was evaluated. A computer was programmed to identify concepts of varying difficulty using a logical elimination strategy. The results indicated that the time (number of steps) required for the computer to identify each concept was inversely related to the structural simplicity of the rule. These results strongly imply that something other than, or in addition to, simple logical elimination is involved in human concept identification strategy.

Furthermore, the hypothesis of differential rule difficulty, even though plausible, still does not seem to account for all the data. For example, Neisser and Weene (1962) reported that subjects seemed to have better verbal understanding of complex rules such as "either/or" than of the more rapidly learned (i.e., "easier") conjunctive rules.

In view of all these arguments, Neisser and Weene (1962) suggest that their data can be better explained in terms of a hierarchical organization. According to these authors, the facilitative effect of learning lower level concepts before learning more complex concepts lies in the fact that to solve rule (A. -B) subjects must learn what (A.) and (-B) mean; following the same reasoning, learning (A. -B) will facilitate learning of (A. -B v (-A. B). It thus appears important to turn to the issue of hierarchical conceptual learning.

Hierarchical Organization of Concepts

Neisser and Weene's data have a theoretical bearing on a conceptual learning model proposed by Gagné, which is fully described in Gagné (1965). Specifically, Gagné's model describes learning as increasing in stages of complexity and difficulty in hierarchical terms. The differential difficulty of concept learning for ostensibly similar concepts, as reported by Neisser and Weene, corresponds well to his theoretical interpretation.

The basic working principle of Gagné's model is the description of learning as a cumulative process. More specifically, he states that "within limitations imposed by growth, behavioral developments result from the cumulative effects of learning" (Gagné, 1968, p. 178).

Since Gagné has been concerned basically with applied research, his work deals with instructional procedures for the teaching of mathematical concepts (Gagné, 1962a, 1962b, 1965, 1966). In these studies he has shown consistently that a complex task can be broken down into its components such that performance in each step of this sequence is dependent on mastery of the previous steps (e.g., Gagné, 1962).

Gagné's model also involves mostly what he calls "rule" or "principle" learning. A rule or principle is basically a concept but is distinguished from the latter in that:

(1) While attainment of a concept can be shown by means of an identificatory response (concrete concept or concept by observation), the rule or principle has to be demonstrated (abstract concept or concept by definition) (Gagné, 1966).

(2) A rule or principle is composed by associations, motor and verbal chains, multiple discriminations, concepts, and simple rules (in the case of complex rules) (Gagné, 1965, 1968).

One of the implications of a rule or principle (as opposed to a concept by observation) is that it is not "learned: but has to be taught (Gagné, 1966). The distinction here seems to relate to the level of abstraction involved in each of these two kinds of concepts. For example, one might expect a subject to learn to identify the radius of a circle even though he is not able to define what the radius of a circle is. The relevant attributes of the concept are all physically contained in the instance and can be isolated, for example, simply by differential reinforcement. A rule or principle, however, requires relational operations that go far beyond the observable properties of the stimuli (e.g., in the principle of "work"). According to Gagné, even in the case where the subject has mastered all the discriminations and concepts involved in the rule, he is not likely to demonstrate the rule if he has not been taught it. Therefore, rule learning as defined by Gagné seems to differ considerably from the process usually studied in psychological research, which deals with what he calls concept by observation.

Moreover, when one thinks of the concepts that constitute a mathematical rule, it is apparent that the hierarchical organization of information becomes an end rather than a means. In the learning of the rule 2N-1, to learn what "-" means is not a facilitatory device but rather a prerequisite (unless, of course, the rule is changed). This notion of hierarchies comprised by prerequisites is recognized by Gagné.

> The hypothesis is proposed that specific transfer from one learning set to another standing above it in the hierarchy will be zero if the lower one cannot be recalled and will range up to 100% if it can be. (Gagné, 1962, p. 358)

There is enough evidence, however, that rules can be learned at any level independently of learning rules from presumably subordinate levels. Haygood and Bourne (1965) and Bourne (1968) have shown consistently that if subjects are given training in discovering rules there is an improvement from problem to problem much like the phenomenon of learning sets described by Harlow (1959). Moreover, Haygood and Bourne's (1965) study also included a condition in which subjects had to learn both a rule and the attributes. Although the performance of this group was considerably poorer than that of the other two groups (rule learning with attributes given and attribute identification with rule given), there is no doubt that subjects did learn the task.

Therefore, although the relationship between Neisser and Weene's results and the work developed by Gagné seem to complement each other, more basic research is needed to clarify some of the problems in hierarchical organization of concepts.

Experiment

The preceding discussion has presented two major theoretical alternatives regarding the learning mechanisms in progressing through hierarchically structured learning sequences. The concept complexity model proposed by Neisser and Weene assumes little "level to level" transfer, but views progress to be more a function of the type and number of rules involved at each step in the hierarchy. However, Gagné, posits the cumulative learning model, and argues that progress at any step is primarily a function of the number and type of prerequisites already mastered (presumably at previous stages).

Since the goals of this research include the development of experimental learning sequences that can be considered suitable analogues to the learning tasks and sequences currently incorporated in individualized curricula, the above theoretical dispute becomes important, i.e., Neisser and Weene's model assumes that step-to-step cumulative component continuity (stimulus specific) in a learning hierarchy is not nearly so important as is type and number of component rules. Their model seems based almost completely on the component analysis of a logical tree structure.

However, Gagné's model argues that the critical feature in hierarchical learning is the cumulative continuity of component stimulus attributes, which leads to the postulation of phenotypical tree structure in hierarchical learning.

To preserve the essential components of both models, the materials to be used in the training and testing of this experiment were designed to provide a crude test of the learning assumptions of each. This design consideration involved the generation of two sets of materials (described later), one based on Gagné's cumulative model and a parallel set based on Neisser and Weene's logical model. One advantage of this strategy (the use of two materials sets) is that it enables for an evaluation of these competing concept learning models. Another advantage is that it protects against the likelihood that the validation evidence for the mastery test model will be atypical or biased with regard to assumptions of hierarchical learning.

A second manipulation that is based on theoretical assumptions pertains to the extent of training required to generate variations in mastery likelihood. This is an important consideration because the mastery model is essentially based on an all-or-none learning assumption, and although individual differences will occur, the range of training will be critical for purposes of establishing these variations in mastery. That is, if the range is too low, too few "mastery" cases will result, and if the range is too high, extremely few cases will be at nonmastery (in addition to the boredom and fatigue factors discussed previously).

Two relevant theoretical models used in developing the training range were Trabasso and Bower's Relevant and Redundant Cue model, (1968), and Estes' Stimulus Sampling Model. Interestingly, although these two models are based on somewhat different assumptions, both lead to nearly identical estimates of the average number of training trials necessary to produce mastery (or attainment) of a concept. Thus, for each concept to be included in the training sequence, the estimate of the average number of trials to mastery was derived and used to define the moderate or middle level training condition. The two other training levels used in this experiment were defined relative to this middle level as follows: low or level 1 was set equal to one-half the number of trials for level 2, high or level 3 was set equal to twice the number of trials for level 2. This procedure yielded the schedule displayed in Table 3 (schedule equals number of trials) for the concept hierarchy used in the experiment.

Table 3

Number of Training Trials by Level
for Each of the Concepts in the Experiment

	Concept Hierarchy		
Training Level	Concept 1	Concept 2	Concept 3
Level 1	10	10	20
Level 2	20	20	40
Level 3	40	40	80

These three training levels effectively operationalize the mastery likelihood parameter of the mastery evaluation model.

The two other factors or parameters described in an earlier section, are test length and item error likelihood. For this experiment, a test corresponds to blank trials, i.e., training trials are characterized by stimulus presentation, response interval, and feedback or knowledge of correct response (KCR). Test trials are characterized as stimulus presentation, response interval, but no feedback or KCR. Test or blank trials are administered following each training sequence.

The test length parameter was operationalized in this experiment at two levels: five and ten items. These levels or lengths correspond roughly to the range of most single skill tests, such as the curriculum imbedded tests of IPI. The item error parameter was operationalized in terms of two item forms for the blank trials (i.e., tests). These were two choice items, corresponding theoretically to true-false tests, and four choice items, corresponding to four option multiple choice items. The assumption of alpha errors being greater for two than four choice items constitutes the item error manipulation.

An additional factor was incorporated in this experiment control for the effects of repeated blank trials. This factor described as test continuity, essentially splits the sample into two groups: one receiving blank trials immediately after each training sequence, and another receiving these trials only after the completion of all training. These two groups are designated as continuous and terminal testing, respectively.

Chapter II

METHOD

Research Design

The experiment consisted of a three-way repeated measures analysis of variance design for the training/transfer manipulation, fully crossed with a three-way analysis of variance design on the test manipulations. These fully crossed factors are as follows:

(1) Training/transfer factors

- Transfer paradigm: cummulative intradimensional attribute shifting (Neisser and Weene) versus no-shift (Gagné)

- Training level: low or level 1 versus moderate or level 2, versus high or level 3, where $L_1 = 1/2\ L_2$, $L_3 = 2 \times L_2$

- Conceptual hierarchy: concept 1 (three dimensions, one relevant) to concept 2 (four dimensions, two relevant) to concept 3 (six dimension, four relevant)

(2) Test factors

- Test length (five item versus ten item)

- Item form (two choice versus four choice)

- Continuity (continuous versus final)

These factorially balanced design factors are presented schematically in Figure 1a for the training components and in Figure 1b for the testing components.

Transfer Paradigm	TNG Level	Concept 1	2	3
	1	n = 16	→	→
No-shift	2	n = 16	→	→
	3	n = 16	→	→
	1	n = 16	→	→
Shift	2	n = 16	→	→
	3	n = 16	→	→

N = 96

Figure 1a Schematic representation of the research design for the training phase of the experiment.

Test Length	Item Type	Test Continuity: Continuous	Final
5-Item	2-Choice	n = 12	n = 12
	4-Choice	n = 12	n = 12
10-Item	2-Choice	n = 12	n = 12
	4-Choice	n = 12	n = 12

N = 96

Figure 1b Schematic representation of the research design for the testing/validation phase of the experiment. This design was fully crossed with the training phase design.

Subjects

A total of 96 third grade boys and girls were recruited for this experiment from three elementary schools in Western Massachusetts. These schools were in East Whately, Greenfield, and West Springfield school districts. Within each school, children were assigned to experimental conditions at random, with the one exception that sex be uniformly distributed across conditions.

Materials

The materials that composed the concepts to be identified in this experiment consisted of brightly colored geometric forms of varying sizes and shapes and on which discernable patterns or textures had been imprinted. These stimuli comprised the conceptual dimensions of shape, color, size, and pattern. Two additional dimensions of position and number were generated through the use of varying numbers of Xs that were situated either above or below the stimulus form.

For a given trial (training or testing), stimuli were arranged in a four choice display, according to a schedule described below, and photographed on a 35mm color film. These photographed trials were mounted in slide frames and presented via slide projector in the training and testing phases of the experiment.

Each child recorded his choice for a given trial by marking a corresponding box or position in his response booklet. For training problems, this booklet consisted of one page for each trial. Each page contained four empty boxes corresponding to the four stimulus positions in the color slide. The child simply marked the box in his answer booklet to indicate which of the four stimuli he chose, and then after knowledge of the correct response (KCR), turned to the next page for the next trial. Test response materials differed in that five two- or four-choice trials were contained on a single page. Examples of training and test response materials are presented in Appendix A.

Conceptual Learning Tasks

As stated above, the training sequence involved the orderly progression of conceptual complexity in hierarchical fashion across the three conceptual attainment tasks. This progression involved the addition of both relevant and irrelevant stimulus dimensions from task to task, and for shift conditions, the changing of "relevant attributes within dimensions. Using the notations of letters representing dimensions and numbers representing attributes, the task sequences are schematicized in Table 4 below (where + = relevant dimensions, - = irrelevant dimensions):

Table 4

Task Sequence for the Concept Attainment Problems

Paradigm	Concept 1	Concept 2	Concept 3
No-shift	+ --	+ + --	+ + + + --
	A_1,BC	A_1,B_1,CD	A_1,B_1,D_1,E_1,CF
Shift	+ --	+ + --	+ + + + --
	A_2,BC	A_3,B_2,CD	A_4,B_3,D_2,E_2,CF

The problems or trials constituting these learning tasks were generated by a computer program, such that both within and across tasks all dimensions and attributes were arranged orthogonally under the restriction that each trial contain one and only one positive instance. Examples of each of these problem sequences (concept by paradigm) and of the test items associated with each are presented in Appendix A. These sample sequences display the problems as presented to the child and the "correct" choice (concept examplar) is designated with a "+".

Apparatus

The training and test materials were projected via a 35 mm carousel slide projector on a screen in full view of the Ss. Also aside from the introduction and pretraining that was presented verbally by the E, all subsequent training and testing instructions were presented via a magnetic tape recorder. The Ss responded individually to each of the training and testing problems by marking their choice in a response booklet.

Procedure

The experiment was conducted in three phases: a short warm-up session, the training phase, and the testing or blank trials phase. The goals of the warm-up sessions were as follows:

(1) Since feedback would be presented via magnetic tape it was decided that confirmation of the correct response would be given in terms of the position occupied by the positive stimulus. Consequently, one of the goals of the warm-up session was to test or teach the understanding of ordinal position, and the ability to match the position of the stimuli with the corresponding spaces provided in the answer sheets.

(2) The second objective of the warm-up session was to acquaint subjects with the ultimate goal of the problem, namely to identify and choose the correct stimulus for each trial, and to discover the conceptual rule.

The training phase consisted of a series of trials with feedback appropriate to the concept to be identified. Testing consisted of blank or nonfeedback trials. The experiment was conducted such that eight children were escorted from their classroom to the experimental room and seated. They were instructed to fill out certain information on the training booklet in front of them. This information included their name, their age, sex, and seat number.

Children then were given some preliminary instruction concerning the nature of the task in which they were to engage. This pretraining included a brief instructional unit in which they were taught how to make responses for specific choices on the screen and also an introduction as to the nature of the specific problems that they would be attempting to solve. Specifically, the children were told that they would be playing a learning game, as follows:

> The nature of the game will be for you to choose the correct picture when I show you several pictures on the screen like this (the slide projector was then turned on and four stimulus figures appeared on the screen). Here we see four pictures. This is the first picture (the E points to the leftmost picture), this is the second picture (E points to the second picture), this is the third picture, and this is the fourth picture (he points to the rightmost picture). Now look at the first page of your booklet. Do you see

those four boxes? (E waits for Ss to acknowledge). Each one of those boxes goes with a picture you see on the screen. The first box would go with the first picture, the second box would go with the second picture, the third box would go with the third picture, and the fourth box would go with the fourth picture. Now, everybody look at the pictures again. Do you see the circle? (Pause)

All right, now suppose that you wanted to choose the circle. Mark an X on your answer sheet that shows that you are choosing the circle. (Pause) How many people chose the third box? Raise your hand if you chose the third box. (Pause) All right, let's try another one. Turn over to the next page. (E then projects a new slide on the screen in which the circle moves to position 2.)

All right, now let's see if you remember how to play this game. Suppose that you wanted to choose the circle again. Mark the box that would show that you are choosing the circle. How many chose the second box? (Pause) Very good. All right, let's try once again. (E advances to a new slide.) Turn to the third page. Now mark the box for the circle. How many marked the first box? (Pause) Very good.

From now on I'll be talking to you over the tape recorder but I want you to keep in mind a few things that are very important. First, this is a learning game so you want to do your best but you also want to be sure that you do your own work. Don't be concerned with what your neighbor is doing because he may be doing things wrong. Second, we'll have a lot of problems to do and each problem goes on a different page. I'll tell you which page it goes on so you be sure you look to see that you are on the correct page. It is very easy to skip a page and be on the wrong one, so look very carefully. Third, once you've made a mark for your choice, don't change it. If you have a problem, simply raise your hand and we'll help you. All right? Very good. I'll be talking to you on the tape recorder from now on. Remember, if you have a problem, just raise your hand.

The rest of the experiment was presented automatically via the magnetic tape recorder and slide projector. Two Es participated in this training, and occasionally a third was added to assist in the training. For the first five or six problems, the second E stood at the front of the room and when the correct choice was announced via the tape recorder he also indicated the correct choice by pointing a marker on the projector screen. The instructions presented on the magnetic tape recorder initially introduced the Ss to the specific nature of the problems they would be solving as follows:

> All right, boys and girls, we're now ready to begin. Now as we explained to you, the purpose of this game is to choose the correct picture. Now, when I show you a problem on the screen, look carefully at each picture. Then, when I tell you, choose one of the pictures by making a mark in your booklet. After everybody has had time to choose the picture, I'll tell you which picture was right so you can see if you chose the correct one. Now, there's a reason why certain pictures are correct and others are not. When you discover this reason, you'll be able to get all of the problems right. So this means that a first you'll get some of the problems wrong. Don't feel bad but try to find the secret so you'll get the rest correct. Work quickly but carefully. Do your own work and don't change any answers once you've made them. I'll say which page each problem goes with so you'll be sure that you're not on the wrong page. All right, let's begin.
>
> Here is the problem for page one. You all should be on the first page of your booklet. See each picture carefully. Now mark the one you think is correct. (Pause) If you marked the third picture, you were correct. The third picture.

This procedure was repeated for each of the training problems. The number of problems presented was determined by the learning condition and the concept level of the particular training sequence.

The instructions given to the children receiving continuous testing (in this case, the first concept tested) are as follows:

> All right, let's continue with the game only we're going to play it a little differently than before. Each of you has a sheet of paper on which you have written your name. Now I'll show you some problems just like before and for each problem you are to choose the picture that you think is correct. However, I'm not going to tell you which one is

correct for these problems. All right, now I'll tell you which line you should be on for each of these problems.

Initial problems were presented at a fairly stable rate of 15 seconds observation time and 10 seconds response interval. For later trials this rate was advanced to roughly 10 seconds observation time and 5 seconds response and feedback time (such that four problems per minute were presented for the later slides in the sequence). The test items were presented at a fairly stable rate of 15 seconds per item; there was no feedback interval.

The eight Ss who served simultaneously at each session of the experiment actually constituted four subgroups of two Ss each. One subgroup of Ss remained throughout all activities for a given training condition, i.e., they received all training and all test items. The second group received only the first five of each ten item test. They were excused from the room and waited outside after they completed the first five items for each of the three tests. The third and fourth groups were excused from the experiment immediately following training for the first and second concepts. They were reintroduced after the tests were completed.

All children received the first five items of the terminal test. However, only the first and third groups of children received the last five items of the terminal test. This procedure did not produce any noticeable negative side effects, particularly with the children who remained throughout the experiment (i.e., received all training and testing). Moreover, the children who did not receive continuous testing (i.e., were excused from the experiment during the first and second tests) appeared concerned that they were not able to participate in everything.

Data Processing

Data in the form of item choices in the training and test booklets were coded and transferred to punched card forms for computer processing. This processing included the evaluation of literal response protocols for the operation of various solution strategies and possible stimulus biases on the part of the Ss. One such bias did appear and was associated with the pretraining and concept 1 problem received by the shift Ss. Specifically, the circle shape used for pretraining corresponded to the trial 1 correct figure for the concept 1 problems under shift training. Subsequent trials tended to correct for this false solution (alpha error) by teaching the _Ss_ that the pattern rather than shape was relevant. However, the initial bias did tend to lead to many early solutions with these _Ss_.

Chapter III

RESULTS AND DISCUSSION

The presentation and discussion of results in this chapter is organized into three major sections. These sections are analysis of training data, analysis of test data, and validation evidence for the test model.

Analysis of Training Data

Response protocols obtained from the training booklets were organized according to the experimental treatment factors for purposes of assessing training effects. The two purposes served by these analyses were: (1) the evaluation of differential learning and transfer effects as predicted by the two conceptual training paradigm (shift versus no-shift) and the establishment of the functional relationships between these training paradigms and other design factors; and (2) the determination of the extent to which differences in performance corresponded to the amount of training provided. This second purpose was principally relevant to validating the assumption that differences in probability of mastery existed and corresponded roughly to the experimental training variables.

The dependent variable selected for this analysis was the number of correct responses out of the last 10 training trials for each of the three experimental concepts. Thus each S contributed three scores (one for each concept) to this analysis. Furthermore, to control for the effects of individual differences resulting from ability differences or premature solutions (i.e., the concept 1 for Shift groups), Ss were blocked into High or Low groups (median split) based on their first 10 trials performance on concept 1. This provided an added factor to the design in the form of a two level covariable block.

The results of a repeated measures analysis of variance performed on these training data are summarized in Table 5. The design factors in this analysis were transfer paradigm (shift versus no-shift) by training level (level 1, level 2, and level 3) by initial performance block (high versus low) across the three concepts trained. Performance

Table 5

Analysis of Variance on Performance (Number Correct of Last 10 Trials) Across Training Factors and Conditions

Source	df	MS	F
Paradigm (P)	1	21.67	1.94
Trng level (L)	2	99.21	8.90**
Ability block (B)	1	166.53	14.94***
P × L	2	2.33	0.21
P × B	1	12.92	1.16
L × B	2	44.88	4.03*
P × L × B	2	17.32	1.55
Error (btwn)	84	11.14	
Concept (C)	2	47.96	11.96***
C × P	2	22.54	5.62**
C × L	4	7.90	1.97
C × B	2	12.41	3.09*
C × P × L	4	2.96	0.74
C × P × B	2	0.67	0.17
C × L × B	4	8.82	2.20*
C × P × L × B	4	14.75	3.68**
Error (within)	168	4.01	

* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$.

means comprising this analysis are presented in Table 6, both in terms of the experimental design and in summary form in terms of design variables.

Significant effects resulting from this analysis are as follows:

(1) Ss in level 2 or level 3 training groups were performing substantially better than Ss in level 1 training. This effect, displayed in Figure 2 provides support for the principal manipulation of the experiment--namely, that the

Table 6

Performance Means (Percent Correct) for Last
10 Trials for Each Concept
In Experimental Training

Design Factor	First Trials (ability)	Concept 1	Concept 2	Concept 3
No-Shift				
L_1	LO	32.5	48.8	40.0
	HI	66.2	65.0	37.5
L_2	LO	51.2	60.0	46.2
	HI	80.0	83.8	88.8
L_3	LO	66.2	52.5	71.2
	HI	78.8	75.0	68.8
Shift				
L	LO	50.0	50.0	32.5
	HI	92.5	70.0	48.8
L	LO	56.2	67.5	67.5
	HI	98.8	85.0	62.5
L	LO	96.2	78.8	57.5
	HI	71.2	68.8	57.5

Summary of Means

Factor	Mean	Factor	Mean
Shift	61.8	LO 1st trials	56.9
No-Shift	67.3	HI 1st trials	72.2
L_1	52.8	Concept 1	70.0
L_2	70.6	Concept 2	67.1
L_3	70.2	Concept 3	56.6

probability of concept mastery was different as a function of training condition. However, as was evident in Figure 2, this effect is concentrated in level 1 versus "other" companions, since the difference between level 2 and level 3 is negligible.

(2) Initial performance differences persist throughout training as evidenced in Figure 3. This effect suggests that individual differences, perhaps in concept identification strategies, are stable across training.

(3) Overall performance systematically declined across the three experimental concepts. This outcome, plotted in Figure 4, is consistent with the assumption that the training concepts are ordered in terms of difficulty. Furthermore, this gradual decline in performance evidenced in Figure 4 would necessarily occur if, as hypothesized prerequisites in terms of previous concepts were either not sufficiently mastered (L_1 training) or overlearned (L_3 training), thus producing negative or "offsetting" transfer.

(4) The above curvilinear training effects interpretation receives further support from inspection of means presented in Figure 5 for the training level by ability grouping interaction. In particular, the low ability groups show a linear performance trend in terms of training level, whereas the trend is curvilinear for the high ability Ss. Since the high groups arrive at a solution earlier within each training sequence, they effectively receive more reinforced practice on the criterial attributes than do the low ability groups. In some instances this would be expected to result in overlearning or a form of functional fixedness, which would interfere with or produce negative transfer for the learning of subsequent concepts.

On the other hand, very few of the low ability Ss appear to reach early solutions or attainments, and therefore would be expected to benefit from extended training. This does appear to be the case.

(5) Further support for this differential transfer interpretation is provided in examining the interactions of: (1) paradigm by concept, (2) ability blocks by concept, (3) training levels by concept by ability block, and (4) paradigm by ability block by training level by concept. Each of these interaction effects is significant and each displayed a pattern

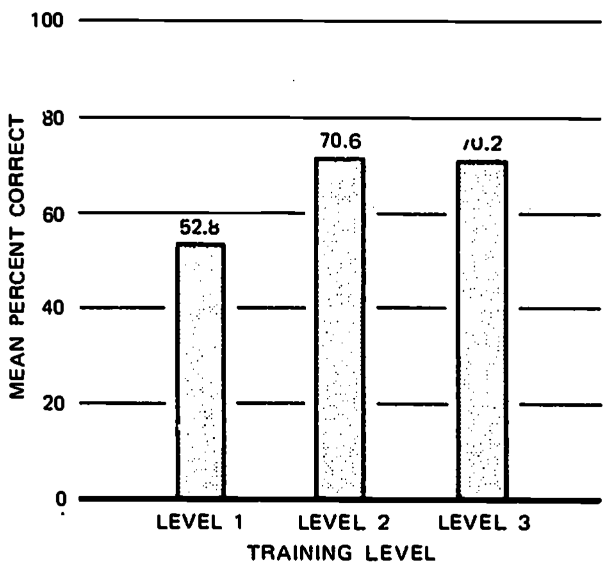

FIGURE 2 MEAN PERFORMANCE (Percent Correct) ON LAST 10 TRAINING TRIALS AS A FUNCTION OF LEVEL OF TRAINING. $F_{(2, 84)} = 8.90$, $P < 0.01$

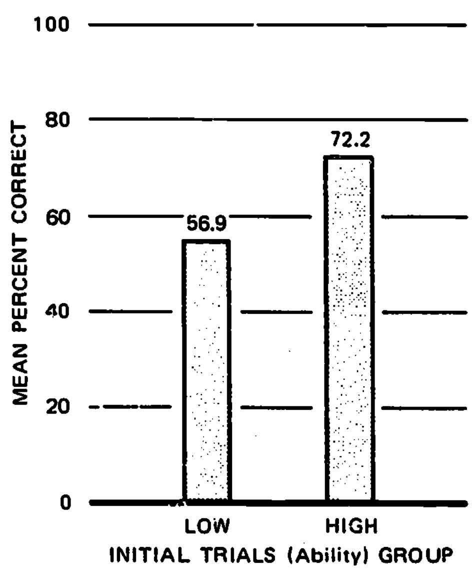

FIGURE 3 MEAN PERFORMANCE ON LAST 10 TRAINING TRIALS AS A FUNCTION OF INITIAL TRIALS PERFORMANCE (Ability Block). $F_{(1, 84)} = 14.94$, $P < 0.001$

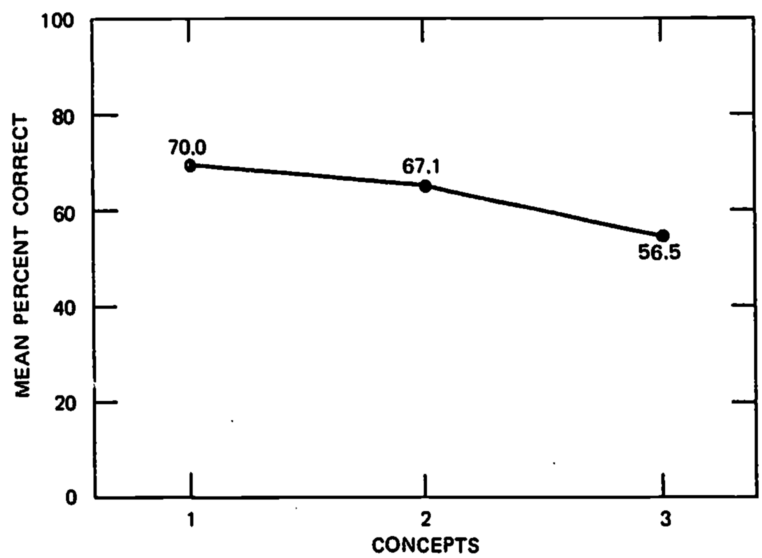

FIGURE 4 MEAN PERFORMANCE ON LAST 10 TRIALS ACROSS THE THREE SEQUENTIAL CONCEPTS. $F_{(2,\ 168)} = 11.96,\ P < 0.001$

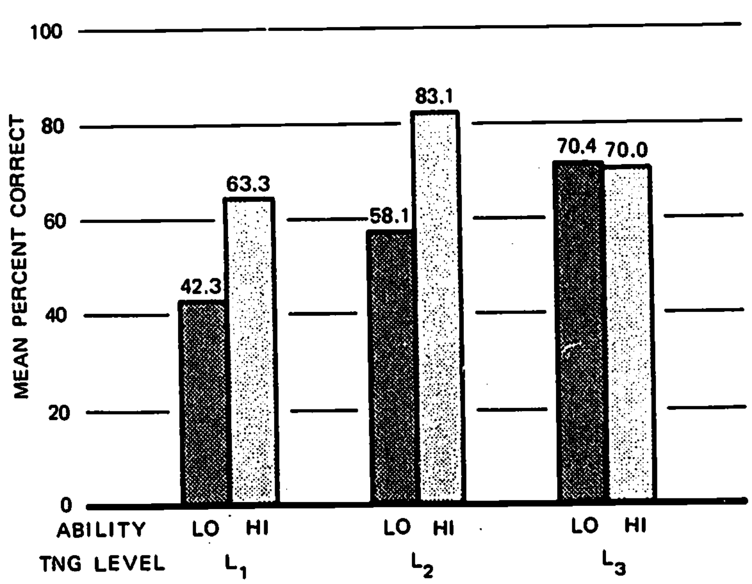

FIGURE 5 PLOT OF DIFFERENTIAL PERFORMANCE IN TRAINING IN TERMS OF THE ABILITY GROUP BY TRAINING LEVEL. $F_{(2,\ 84)} = 4.03,\ P < 0.05$

consistent with the above outlined interpretation. These interactions are presented graphically in Figures 6, 7, 8, 9. Specifically, the overall decrement in performance was greater for concepts involving intradimensional shifts than for the no-shift paradigm (Figure 6). Furthermore, Figure 6 indicates essentially no decrement in the no-shift condition. Figure 7 demonstrates a similar pattern with regard to initial performance groups. The high groups show the decrement across problems, whereas the low groups remain relatively stable. Both groups, however, are significantly above chance across all the problems.

Figure 8 displays this differential performance--or transfer pattern--in terms of training level for each ability block. This effect appears concentrated in the shift from concept 1 to concept 2 in that for the low initial groups, level 3 training groups experienced a decrement whereas level 1 and level 2 groups tended to improve. However, the high initial groups all show declines across the three concept problems, again suggesting possible negative transfer (or--at least--a return to baseline performance).

Finally, Figure 9 shows these transfer effects to be substantially different for the two paradigms, i.e., performance tended to decline systematically for high block shift groups, regardless of training level, whereas moderate (L_2) training appeared facilitating for the low block groups. Effects appear negligible, if not slightly positive for the no-shift training method from concept 1 to concept 2 (except for the level 3 training) and appear erratic for concept 2 to concept 3 transfer.

It should be noted that the interpretations of transfer as applied to the present evidence is somewhat unorthodox. The general transfer paradigm of:

		Train		Test
	Group 1	A	⟶	B
vs.				
	Group 2	rest		B

is not represented in this analysis, since the principal focus of this research was not that of experimenting on transfer. However, prior

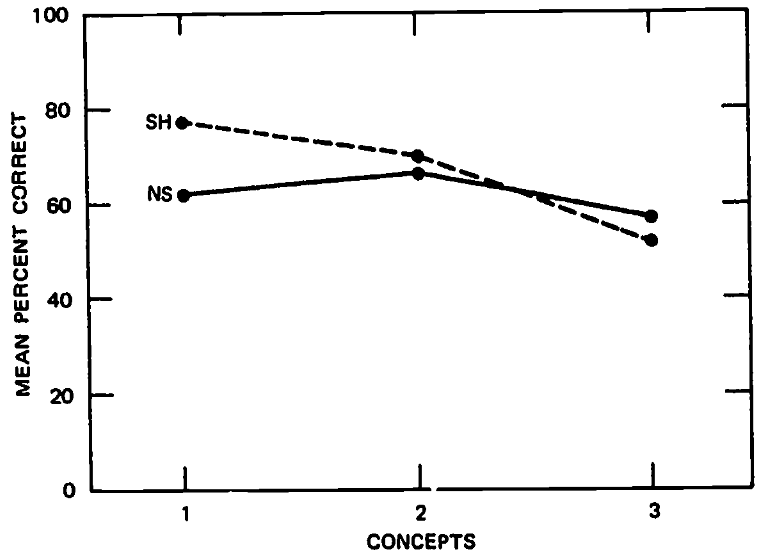

FIGURE 6 PLOT OF THE DIFFERENTIAL EFFECTS OF THE TWO TRANSFER PARADIGMS (Shift versus no-shift) ON TRAINING PERFORMANCE ACROSS THE CONCEPTUAL HIERARCHY. $F_{(2,\ 168)} = 5.62$, $P < 0.01$

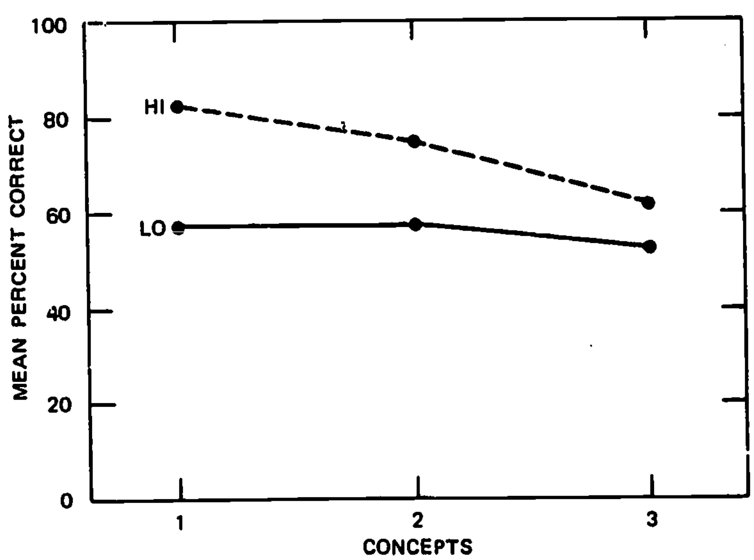

FIGURE 7 PLOT OF THE DIFFERENT TRAINING CURVES FOR EACH OF THE TWO ABILITY GROUPS ACROSS THE CONCEPTUAL HIERARCHY. $F_{(2,\ 168)} = 309$, $P < 0.05$

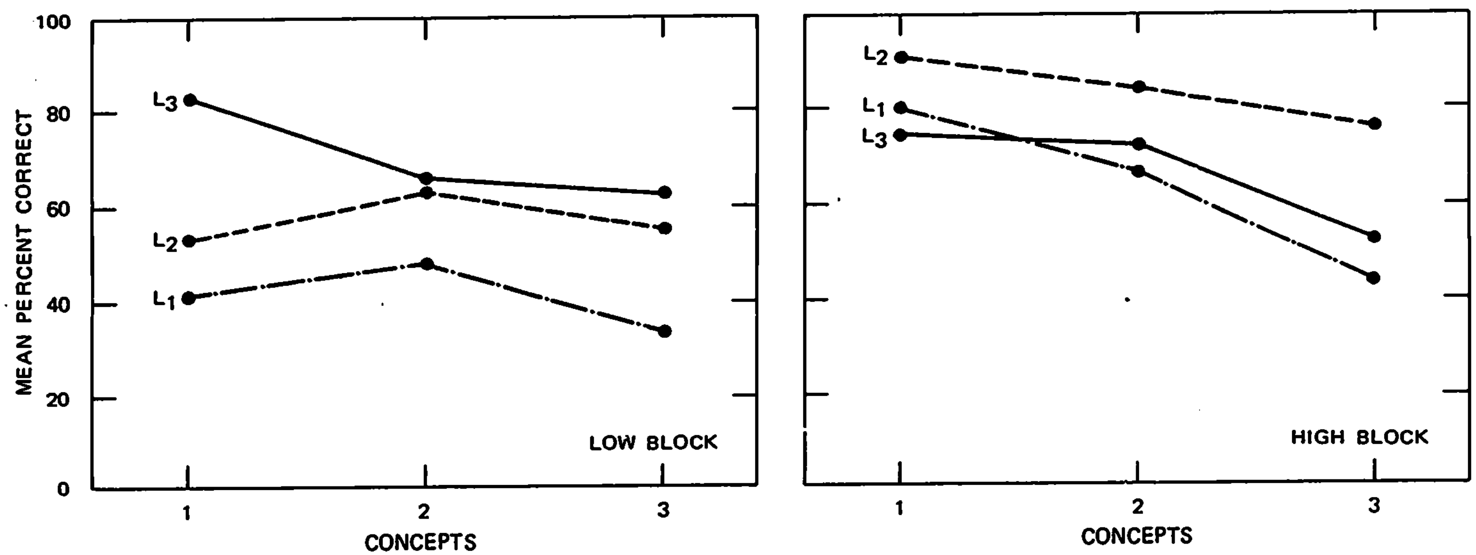

FIGURE 8 PLOT OF THE SEPARATE DIFFERENTIAL EFFECTS OF LEVELS OF TRAINING ACROSS THE CONCEPT HIERARCHY FOR EACH OF THE INITIAL TRIALS ABILITY GROUPS. $F_{(4,\ 168)} = 2.25,\ P < 0.05$

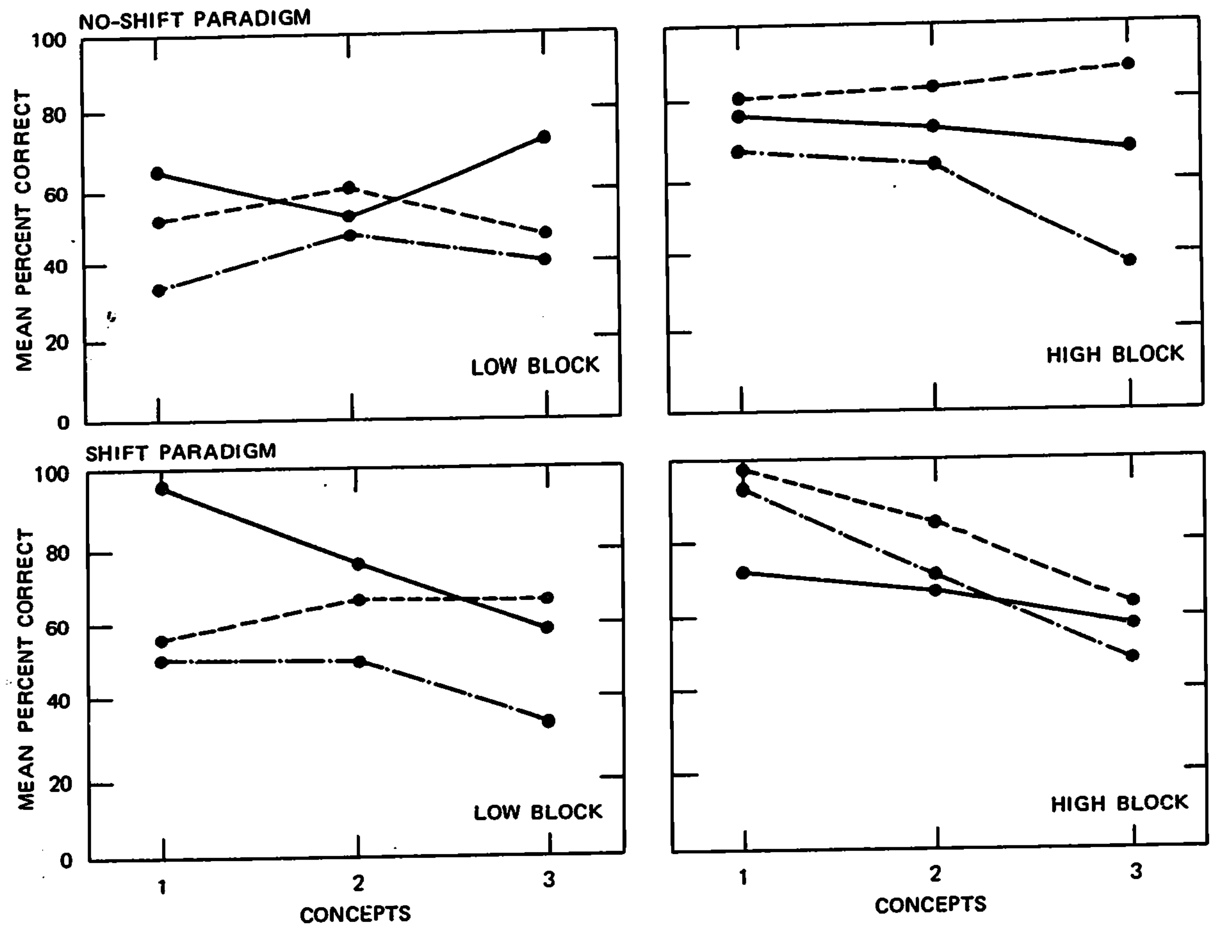

FIGURE 9 PLOTS OF THE DIFFERENTIAL EFFECTS OF TRAINING LEVEL ON PERFORMANCE ACROSS THE CONCEPT HIERARCHY SEPARATELY FOR INITIAL TRIALS GROUPINGS WITHIN EACH TRAINING PARADIGM. $F_{(4,\ 168)} = 3.68,\ P < 0.01$

research on transfer parameters has provided the proposed interpretation that appears to fit--or, at least, is not inconsistent with--these training data.

To some extent, the training data can stand alone in providing evidence for the validity of the assumed manipulations. The no-shift concepts were responded to in a consistent fashion, and, correcting for initial solution bias of the shift training, with fewer errors. The levels manipulation evidenced its effects throughout the data, and particularly for the concept 3 solution, which was clearly the most difficult to attain. However, at least three sources of experimental error were present in these training data. These sources are: (1) the possibility of Ss skipping pages accidentally, (2) of waiting until feedback--or KCR--before entering their response, and (3) of changing an entry following KCR. Thus, it is necessary to analyze test data to establish and corroborate these experimental effects.

Analysis of Test Data

Two separate analyses of variance on test data (blank trials) were performed; the first analysis was formally identical to the analysis performed on the training data, with the exception of the omission of the blocking factor. This analysis was performed on all those Ss receiving blank trials after each training series (i.e., one-half the sample). The second analysis was performed on the total sample but for blank trial performance on concept 3 problems only. The results of these two analyses are presented and discussed separately.

For each analysis, the dependent variable was the percent correct on the blank trials. Also, since the testing factors (item form and test length) are analyzed separately in the model validation section, they were not included in the analyses of variance. The results of the analysis of test performance across concepts are summarized in Table 7, and the cell means are presented and summarized in Table 8. These results are described and interpreted as follows:

(1) Mean performance varied significantly across the three concept tests. Overall performance was best (most correct) on concept 2 and poorest on concept 3 blank trials. This performance differential was relatively uniform, with about 8 percentage points separating one average from the next. However, the trend does not parallel that observed for training data, in that concept 1 and concept 2 performance means are reversed.

Table 7

Analysis of Variance of Performance
(Number Correct) During Blank
Trials Across Concepts 1, 2, and 3

Source	df	MS	F
Paradigm (P)	1	57.51	2.83
Level (L)	2	57.33	2.82
P × L	2	35.19	1.73
Error (btwn)	42	20.33	
Concept (C)	2	25.65	4.68**
P × C	2	50.13	8.81**
L × C	4	9.38	1.65
P × L × C	4	7.54	1.32
Error (within)	84	5.69	

** $p < 0.01$.

Table 8

Performance Means (Percent Correct) for Blank Trials (Tests) for Each Concept

Design Factor		Concept 1	Concept 2	Concept 3
No-shift	L_1	36.2	46.2	31.2
	L_2	62.5	66.2	71.2
	L_3	33.8	75.0	73.7
Shift	L_1	65.0	73.8	46.2
	L_2	75.0	61.2	42.5
	L_3	90.0	93.8	62.5

Summary

Factor	Mean
No-shift	55.1
Shift	67.8
L_1	49.8
L_2	63.1
L_3	71.4
C_1	60.4
C_2	69.4
C_3	54.6

(2) The above outcome becomes more complex when the interaction of paradigm by concept is evaluated. Specifically, as demonstrated in Figure 10, mean test performance is seen to increase from concept 1 to concept 2 for the no-shift groups, whereas it displays a slight decline from concept 1 to concept 2 for shift Ss. The concept 2 to concept 3 differential is disordinal (crossover) in that by concept 3, no-shift Ss are averaging more correct responses than are shift Ss. Thus, based on overall trends, the no-shift Ss tend to show improvement across tests, whereas the shift Ss tend to decline. This outcome is consistent with the training data and with the transfer interpretation proposed earlier.

The results of an analysis of the total sample for concept 3 blank trial performance (recall that only one-half of the Ss received blank trials across concepts 1, 2, and 3) are presented in Table 9, and the corresponding cell means are presented and summarized in Table 10. These results show the single significant performance difference on concept 3 blank trials to be that corresponding to training levels. In particular, level 1 Ss averaged 38.8 percent correct, level 2 averaged 46.2 percent, and level 3 averaged 62.6 percent. These performance averages correspond well to the training manipulations, and likely represent a less biased (methodologically) estimate of the experimental effects.

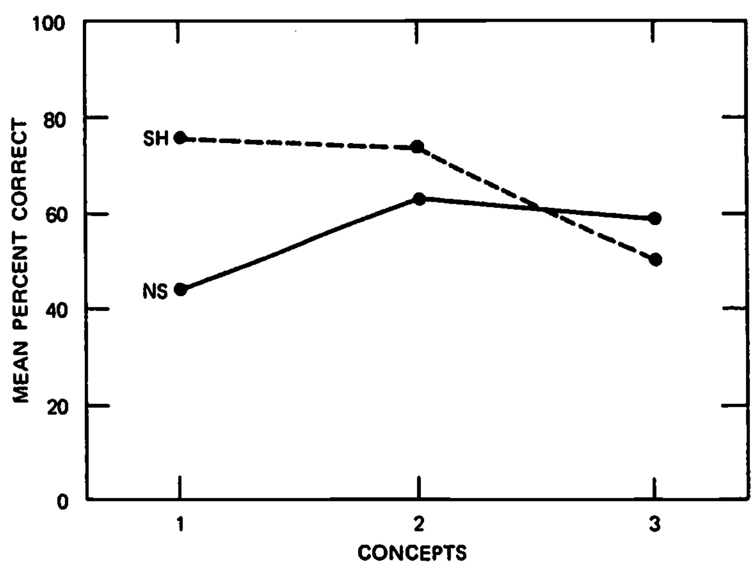

FIGURE 10 MEAN BLANK TRIAL PERFORMANCE (Percent Correct) ACROSS CONCEPTS FOR EACH TRANSFER PARADIGM. $F_{(2,\ 84)} = 8.81$, $P < 0.01$

Table 9

Analysis of Variance on Performance (Number Correct) For Total Sample On Concept 3 Blank Trials

Source	df	MS	F
Paradigm (P)	1	10.01	0.
Level (L)	2	43.26	4.22*
Group (G)	1	31.51	3.07
P X L	2	19.01	1.85
P X G	1	0.84	0.08
L X G	2	9.20	0.90
P X L X G	2	10.41	1.02
Error	84	10.26	

* $p < 0.05$

Table 10

Performance Means (Percent Correct) on Concept 3 Test: Total Sample

		Training Level		
		L_1	L_2	L_3
No-shift	CONT	31.2	71.2	73.8
	FNL	35.0	35.0	66.2
Shift	CONT	46.2	42.5	62.5
	FNL	42.5	36.2	43.8

Summary of Means

Factor	Mean	Factor	Mean
No-shift	52.1	CONT	54.6
Shift	45.6	FNL	43.1
L_1	38.8	C_1	60.4
L_2	46.2	C_2	69.4
L_3	62.6	C_3	52.5

Other outcome trends evidenced in Table 10 are a tendency for no-shift groups to perform better than shift groups, and a tendency for groups receiving blank trials throughout the experiment (continuous testing) to perform better than those introduced only at concept 3. The first of these trends appears consistent with the transfer interpretation proposed in the previous section (training effects), and the second trend suggests a possible familiarization (with blank trials procedures) effect. The important point is that both of these trends, and the preceding significant effects, are consistent with the results of the training analyses and thus provide evidence for the validity of the experimental manipulations. To that extent, the data do appear appropriate for the evaluation of the mastery test model, which is the principal focus of this experiment.

Validation Evidence for the Test Model

The previous two analyses sections evaluated for training and test effects in terms of design variables, such as level of training, paradigm, and initial performance across concepts. The analyses reported in this section will deal with the establishment of test parameters (where a test consists of blank trials) and of the operating characteristics of the mastery test model. The design factors for the validation analyses are paradigm (shift versus no-shift) by test length (five versus ten items) by item form (two versus four choice), across the three conceptual problems.

Three steps were followed in developing evidence for the validity of the mastery model: (1) performing item analysis and subsequently estimating single item reliability for data obtained under each of the test grouping conditions for each conceptual test--this amounted to performing 24 separate item analyses (two test lengths by two item forms by two paradigms by three concepts equals 24); (2) generating optimal cut rules (pass/fail) for each of these "tests" through implementation of the mastery model using parameters generated in (1), above; and (3) estimating the concurrence of mastery/nonmastery decisions based on test data to training data, and evaluating these correspondences in terms of design factors (level of training and so forth) and of goodness of fit for overall test distributions.

Items Analyses

The results of each of the 24 item analyses are presented in Appendix B. For each analysis, subject by item responses are presented, as are total scores and item difficulties (percent passing the item). Test statistics presented for each such analysis are the mean, standard deviation, reliability (as estimated by Kuder Richardson formula 20), and the average item reliability (as estimated by the Spearman Brown formula). These means, standard deviations, and reliabilities are summarized in Table 11.

Inspection of this table fails to reveal any clearly systematic patterns across concepts, particularly in terms of overall test reliability. The obtained values, however, are generally high and acceptable, particularly for such "short" tests. Also the ten item tests do appear to yield, on the average, higher reliability values than do the five item versions. But the expected result of increase in reliability with increase in test length does not consistently occur. The two exceptions both occur with two choice tests and thus may be due to chance

Table 11

Summary of Means, Standard Deviations, and Reliabilities from Test Item Analyses

Design Factor	Concept 1 $\bar{X}$	Concept 1 R	Concept 1 SD	Concept 2 $\bar{X}$	Concept 2 R	Concept 2 SD	Concept 3 $\bar{X}$	Concept 3 R	Concept 3 SD
No-shift									
5 Item-2CH	3.50	.851	1.643	4.17	.716	1.169	3.69	.391	1.155
5 Item-4CH	1.00	.686	1.265	2.67	.543	1.366	1.42	.899	1.881
10 Item-2CH	5.50	.950	3.987	6.00	.564	2.098	6.58	.671	2.275
10 Item-4CH	3.17	.899	3.125	5.33	.880	3.266	4.08	.984	4.562
Shift									
5 Item-2CH	4.67	--	0.516	4.00	.440	1.095	2.08	.651	1.564
5 Item-4CH	3.00	.970	2.280	4.00	.642	1.265	2.67	.670	1.557
10 Item-2CH	8.00	.938	3.098	7.00	.981	4.147	5.92	.326	1.782
10 Item-4CH	7.33	.988	4.131	7.50	.958	3.507	2.83	.820	2.725

effects associated with tests of this type (i.e., the true-false test analogue). In all, the results of this analysis were considered acceptable for purposes of generating mastery cut rules.

Mastery Cut Rules

Three parameters are necessary to compute the optimal mastery cut rule using the formula and model described earlier in this report. These parameters are the length of the test, the item error probability and corresponding class (false pass versus false fail likelihood), and the prior probability of mastery (also described as the relative decision error weights). The test length parameter thus corresponds to five versus ten item tests. Item errors are estimated from total test reliability and are distributed in terms of item form; prior probability of mastery is seen to correspond to training level (L_1, L_2, or L_3).

Using the estimates of average item reliabilities provided by the item analyses, a matrix of cut rules was computed for each of the 24 test item groups. These matrixes, presented below each corresponding item analysis in Appendix B, provide percent correct cutoffs and number correct values required for a mastery decision for each of several prior mastery likelihoods by each of several relative item error weights. The prior mastery likelihoods are 1:100, 10:1, equal, 1:10, and 1:100. Alpha (false pass) to beta (false fail) item weights--for a given item error likelihood, as estimated from the preceding item analysis--are varied as follows: 10:1, 5:1, 3:1, 2:1, 1:1, 1:2, 1:3, 1:5, 1:10. Corresponding cut rules then are listed for these 45 combinations (i.e., the five prior probabilities or ERR WT by the nine relative item error weights or alpha to beta ratio).

The cut rules were applied to the test data as follows: training level was considered equivalent to prior probability of mastery (ERR WT) such that $L_1 = 0.01$, $L_2 = 1.0$, and $L_3 = 10.0$. Item form was considered equivalent to relative item error (alpha to beta ratio) such that two choice = 0.330 and four choice = 0.500. For each test, each score was evaluated in terms of training factors (L_1, L_2, L_3), and item form (two or four choice), and those scores that did not equal or exceed the derived mastery value were interpreted as reflecting nonmastery. This procedure was followed for each of the 24 tests as presented in Appendix B. The specific range of cut rules that were applied to the corresponding test scores are enclosed by the box within each matrix.

To demonstrate the operating characteristics of this evaluation model, overall test data were aggregated into two distributions, one for each test length (five versus ten items). Each of these distributions is comprised of both mastery and nonmastery scores. Application of cut rules as generated by the model on a score by score basis operates to differentiate the two component distributions from the overall distribution. This differentiation is shown in Figure 11 for the five

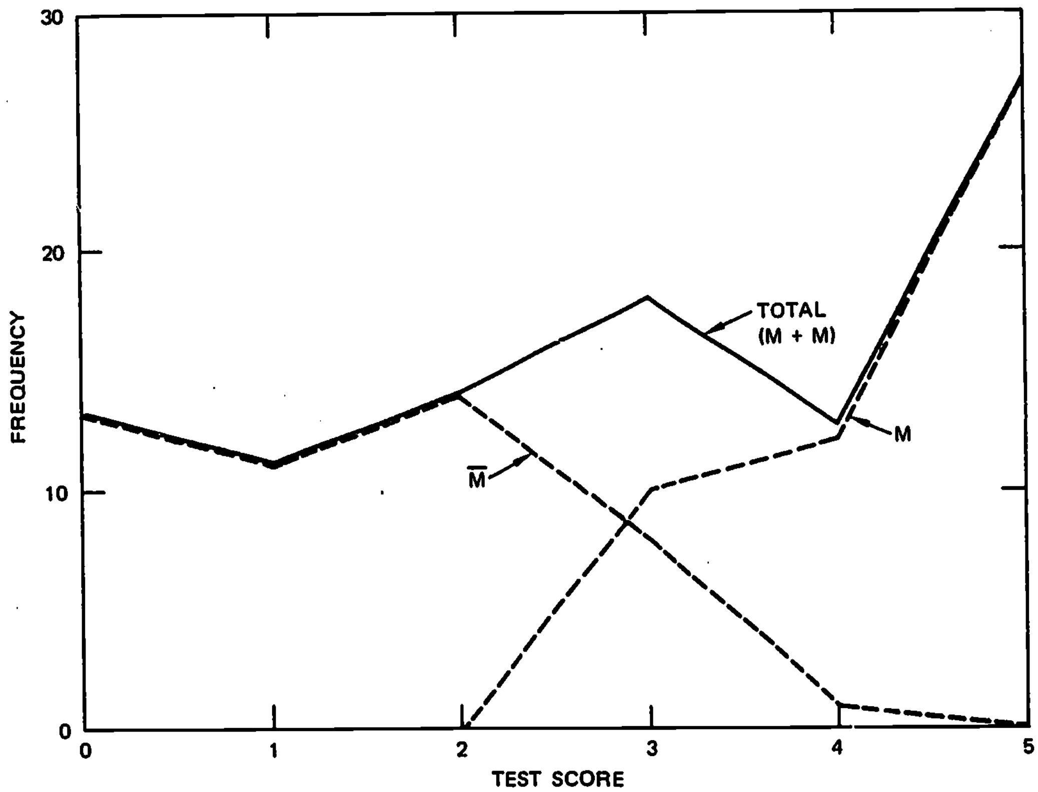

FIGURE 11 DISTRIBUTION OF SCORES ON FIVE ITEM TESTS FOR THE TOTAL SAMPLE (M+M Groups) AND FOR THE INDIVIDUAL M AND M GROUPS AS EVALUATED BY THE MASTERY MODEL

item tests, and in Figure 12 for the ten item tests. In both of these figures, it is clear that the decisions based on the model--applied case by case--are quire different than those that would result from the application of a single across-the-board cut rule. Furthermore, in the case of the ten item distribution, substantial overlap is evident for mastery and nonmastery distributions.

To establish the concurrent validity of these evaluations, test cut rules were applied to corresponding training data (last five trials for five item tests, last ten trials for ten item tests) and evaluated for concurrence of "fit" using chi square procedures. These analyses were performed across all design factors, and separating for item type, conceptual test, training level, and for level by item type. The results of these analyses are summarized in Table 12 and are interpreted as supporting both the conclusions drawn from training and test analyses, and those regarding the validity and utility of the evaluation model.

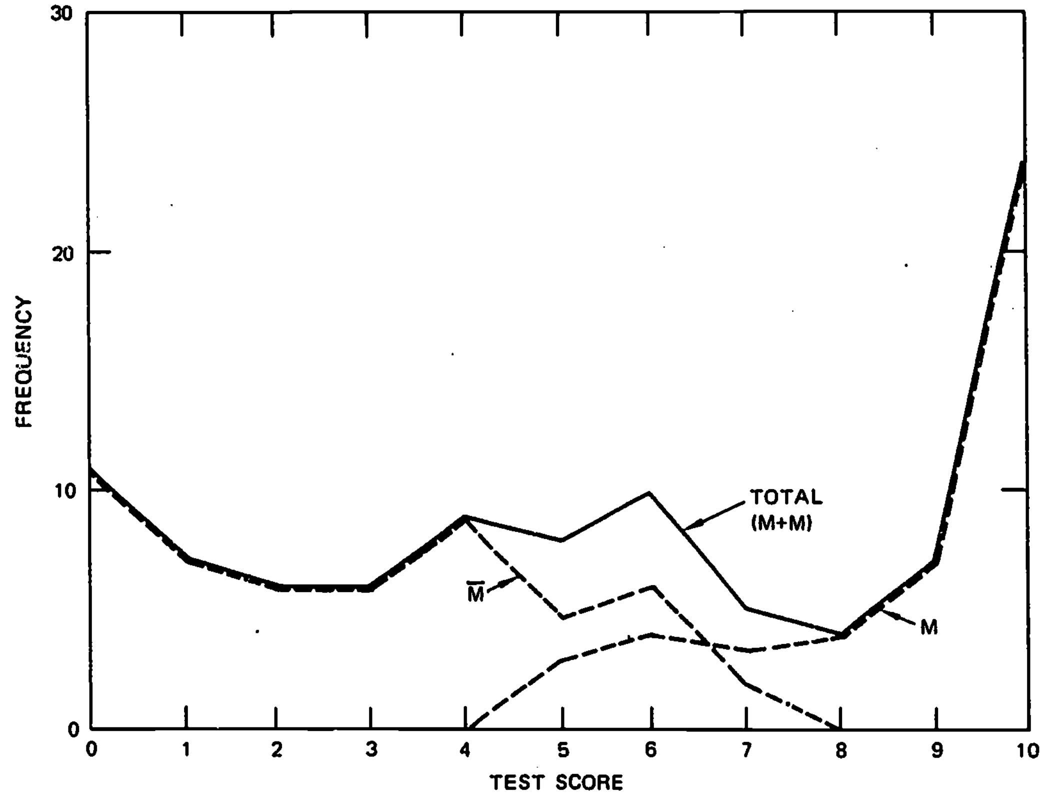

FIGURE 12 DISTRIBUTION OF SCORES ON 10 ITEM TESTS FOR THE TOTAL SAMPLE AND FOR INDIVIDUAL M AND M̄ GROUPS AS EVALUATED BY THE MASTERY MODEL

Specifically, the overall training-testing contingency (χ^2 = 31.51) establishes the upper limit for evidence of validity that can be based on concurrence of training-testing mastery decisions. Since the cut rule formula is optimal in terms of the reliability of the test involved, one estimate of the concurrent validity of the model is the extent to which the training-test contingency corresponds to the mean test item reliability. Expressing this estimate as a ratio of train-test contingency to mean test item reliability, the apparent concurrent validity of the test model is 0.375/0.395 = 0.949. In other words, given the unreliability of the tests involved and the assumption of training testing correspondence, the evaluation model appears 95 percent effective in optimizing test information.

Other evidence for the validity of the model is drawn from similar correspondence of train-test contingencies as a function design factor. For example, the model appears essentially equally effective in differentiation mastery states using two choice (χ^2 = 18.52) and four choice (χ^2 = 20.15) tests. This is particularly impressive given that

Table 12

CHI Square Values for Concurrence of Mastery Evaluations of Training and Test Status, Based on Test Data as Criteria

Level of Analysis		N	χ^2	P
Overall		192	31.51	<0.001
2-Choice tests		96	18.51	<0.001
4-Choice tests		96	20.15	<0.001
Concept 1 tests		48	19.85	<0.001
Concept 2 tests		48	19.93	<0.001
Concept 3 tests		96	11.09	<0.001
Level 1 training		64	5.99	<0.02
Level 2 training		64	7.42	<0.01
Level 3 training		64	11.00	<0.001
2-Choice,	L_1 TRNG	32	3.56	0.06
	L_2 TRNG	32	2.38	0.15
	L_3 TRNG	32	3.41	0.07
4-Choice	L_1 TRNG	32	1.74	0.20
	L_2 TRNG	32	4.57	<0.05
	L_3 TRNG	32	9.50	<0.01

conventionally these tests produce quite different results. Thus, the model appears capable of dealing with the item-error parameter (α/β) quite effectively. Also, the training test correspondence patterns nicely in terms of training level. That is, a systematic transition of cases from the fail-fail to pass-pass contingencies occurs from L_1 to L_3 data.

Finally, goodness of fit tests were applied to the mastery and nonmastery distributions separately for five and ten item tests. The theoretical distributions against which each of these empirical curves were compared are probability distributions of the form:

$$P(n|s) = {}^{n}_{n,t-n}C \; p^{n} \times q^{t-n}$$

where

n = test score (number correct)
C = binonial coefficient
s = mastery state
t = test length
p = item error probability for state S
q = 1 - P

The results of these tests were favorable for the five item distributions, showing reasonably good theoretical and empirical correspondence for mastery ($\chi^2 = 12.8$, $p = 0.05$), and very good fit for nonmastery ($\chi^2 = 5.99$, $0.6 > p > 0.5$) distributions. The ten item distributions, however, showed rather poor fit to the expected curves mastery $\chi^2 = 46.25$, $p < 0.01$ and nonmastery $\chi^2 = 74.83$, $p < 0.01$. One likely reason is that these ten item distributions are multinomial, or at least composits of two or more binomial distributions. As such, they might comprise the basis of subsequent tests of fit in replicate analyses.

Chapter 4

CONCLUSIONS

The results of the analysis of data obtained from this experiment provide several types of evidence that bear directly on the validity and utility of the proposed mastery evaluation model. The first of these classes of evidence pertains to the assumptions of a cumulative and hierarchical learning model--similar to that proposed by Gagné and incorporated in individualized instructional systems such as IPI. Support for these assumptions is provided from the analysis results on training data, in which a curvilinear transfer effect appears to occur for shift and not for the no-shift. The effect becomes more dramatic when controlling for premature solutions such that overall, the data strongly favor the cumulative hierarchical model.

Analysis of training and test performance also supplies strong corroborative support for the assumption that variations in training experience around the expected requirements--as derived from the stimulus sampling model (Estes) or the relevant and redundant cue model (Trabasso and Bower)--effectively produce systematic and operational differences in likelihood of concept attainment (mastery). This outcome is important, since it established an empirical basis for the subsequent validation of the test mastery evaluation model.

The evidence derived in support of this model, although not striking or dramatic, is nonetheless favorable. It was shown that the evaluation model was optimal to the extent that it was 95 percent effective in matching test performance with mastery state, given the constraints implied by the training-testing contingency. It is also concluded that the theoretical and actual test distributions show reasonably good concurrence for the short (five item) tests, but that the fit is not so good for the longer tests. This outcome is favorable at least in part, since the model is designed primarily for use with extremely brief tests.

Therefore, it is concluded that the evidence obtained from this research is supportive of the assumptions of the mastery evaluation model with respect to single skill mastery/nonmastery decisions. To further establish the validity of this model, research should be undertaken in which content valid curricula constitute the material being taught and in which single skill criterion referenced tests, similar to IPI "curriculum imbedded tests," constitute the measuring instruments. Such a study would incorporate branching and tracking, or path analysis of children subsequent to sequential mastery/non-mastery decisions. A study of this nature could be used both to further establish the apparent validity of the model, and to more closely appraise its characteristics on a cost-benefit basis.

REFERENCES

Archer, E. J., Bourne, L. E., Jr. & Brown, F. G. Concept identification as a function of irrelevant information as instructions. Journal of Experimental Psychology, 1955, 49, 153-154.

Battig, W. F. & Bourne, L. E., Jr. Concept identification as a function of intra- and inter-dimensional variability. Journal of Experimental Psychology, 1961, 61, 4, 329-333.

Block, J. H. Mastery learning; theory and practice. New York: Holt, Rinehart, and Winston, 1971.

Bormuth, J. R. On the theory of achievement test items. Chicago, U. of Chicago Press, 1970.

Bourne, L. E., Jr. & Haygood, R. C. The role of stimulus redundancy in concept identification. Journal of Experimental Psychology, 1959, 58, 232-238.

Bourne, L. E., Jr. Concept attainment. In Theodore R. Dixon & David L. Horton (Eds.), Verbal behavior and general behavior theory. New York: Prentice-Hall, Inc., 1968, chap. 10.

Bourne, L. E., Jr. & Guy, D. E. Learning conceptual rules: I. Some interrule transfer effects. Journal of Experimental Psychology, 1968, 70, 3, 423-429.

Bruner, J. S., Goodnow, J. J. & Austin, G. A. A study of thinking. New York: Wiley & Sons, Inc., 1956.

Chernoff, H., & Moses, L. E. Elementary decision theory. New York: Wiley, 1959.

Cooley, W. W., & Glaser, R. An information system for individually prescribed instruction. In: Atkinson, R. C., & Wilson, H. A. (Ed.) Computer assisted instruction. New York: Academic Press, 1969.

Emrick, J. A., & Adams, E. N. An evaluation model for individualized instruction. Research Report RC2674, IBM Thomas J. Watson Research Center; Yorktown Hts., N. Y., October, 1969.

Emrick, J. A. An evaluation model for mastery testing. Journal of Educational Measurement 1971, 8(4), 321-326.

Emrick, J. A. The validation of an evaluation model for individualized instruction. USOE Research Grant, OEG 1-71-0002 (509), September 1970.

Gagné, R. The acquisition of knowledge. Psychological Review, 1962, 69, 355-365.

Gagné, R. M., Mayor, J. R., Garsterns, H. L. & Paradise, N. E. Factors in acquiring knowledge of a mathematical task. Psychological Monographs, 1962b. 76 (7, Whole No. 526).

Gagné, R. M. The conditions of learning. New York: Holt, Rinehart & Winston, 1965.

Gagné, R. M. The learning of principles. In Herbert Klausmeier & Chester W. Harris (Eds.), Analysis of concept learning. New York: Academic Press, 1966.

Gagné, R. M. Contributions of learning to human development. Psychological Review, 1968, 75, 3, 177-191.

Glaser, R. Adapting the elementary school curriculum to individual performance. Proceeding of the 1967 invitational conference on testing problems. Princeton: ETS, 1968.

Ghiselli, E. E. Theory of psychological measurement. New York: McGraw Hill, 1964.

Harlow, H. Learning set and error factor theory. In S. Koch (Ed.), Psychology, A study of a science, vol. 2, New York: McGraw-Hill, 1959.

Haygood, R. C. & Bourne, L. E., Jr. Attribute- and rule-learning aspects of conceptual behavior. Psychological Review, 1955, 72, 3, 175-195.

Lord, F. M., & Novick, M. R. Statistical theories of mental test scores. Massachusetts: Addison-Wesley, 1968.

Neisser, U. & Weene, P. Hierarchies in concept attainment. Journal of Experimental Psychology, 1962, 64, 6, 640-645.

Trabasso, T. & Bower, G. H. Attention in learning: theory and research. New York: John Wiley & Sons, Inc., 1968.

Walker, C. M. & Bourne, L. E., Jr. The identification of concepts as a function of amounts of irrelevant and relevant information. American Journal of Psychology, 1961, 74, 410-417.

Appendix A

SAMPLE TRAINING SEQUENCES AND RESPONSE MATERIALS

Appendix A

SAMPLE TRAINING SEQUENCES AND RESPONSE MATERIALS

62

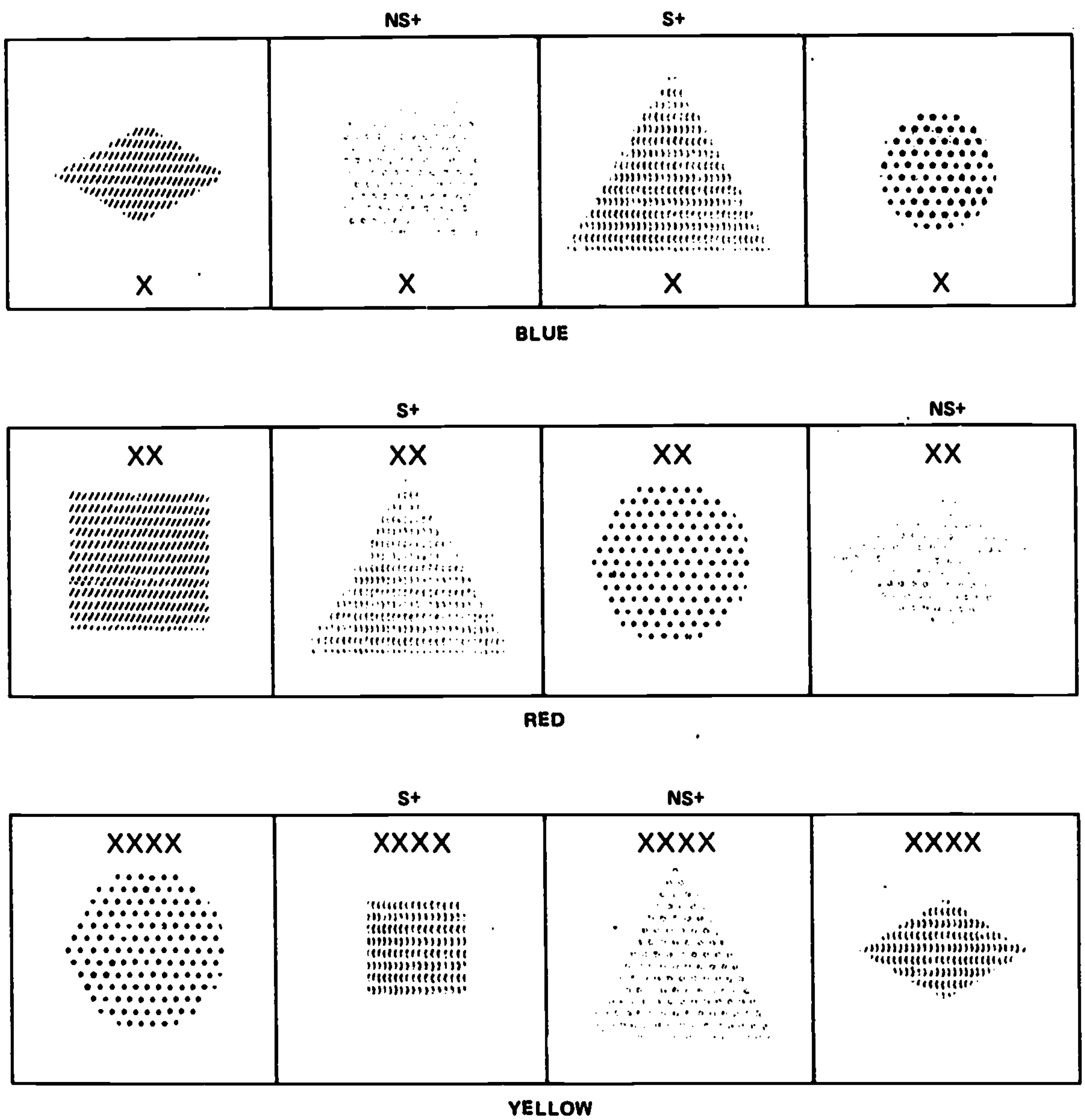

EXHIBIT 1 SAMPLE SEQUENCE OF TRAINING TRIALS FOR CONCEPT 1. THESE STIMULI WERE USED BOTH FOR SHIFT AND NO-SHIFT GROUPS. (S+ = Shift Key, NS+ = No-Shift Key)

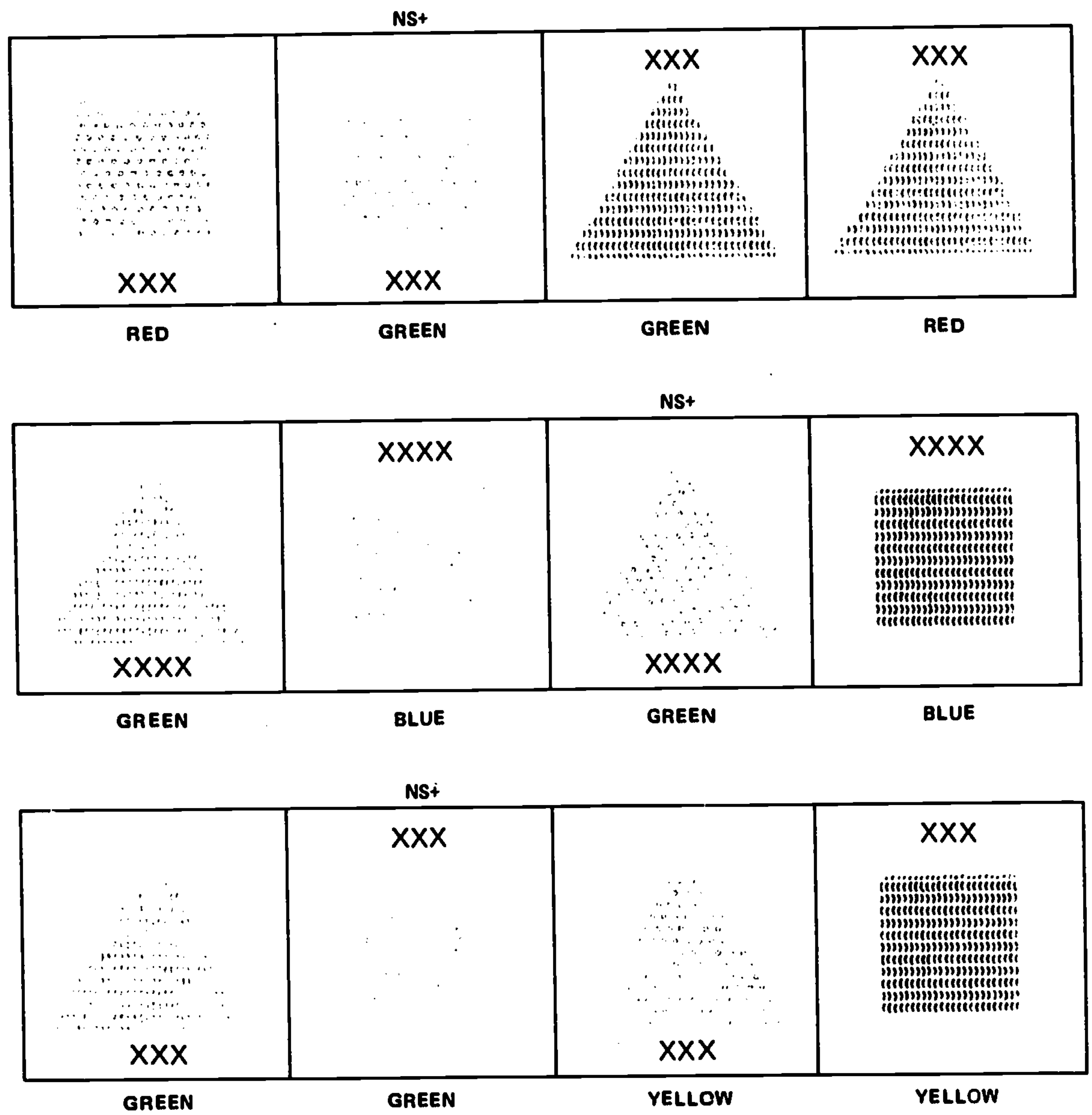

EXHIBIT 2 SAMPLE TRAINING ITEMS FOR NO-SHIFT, CONCEPT 2

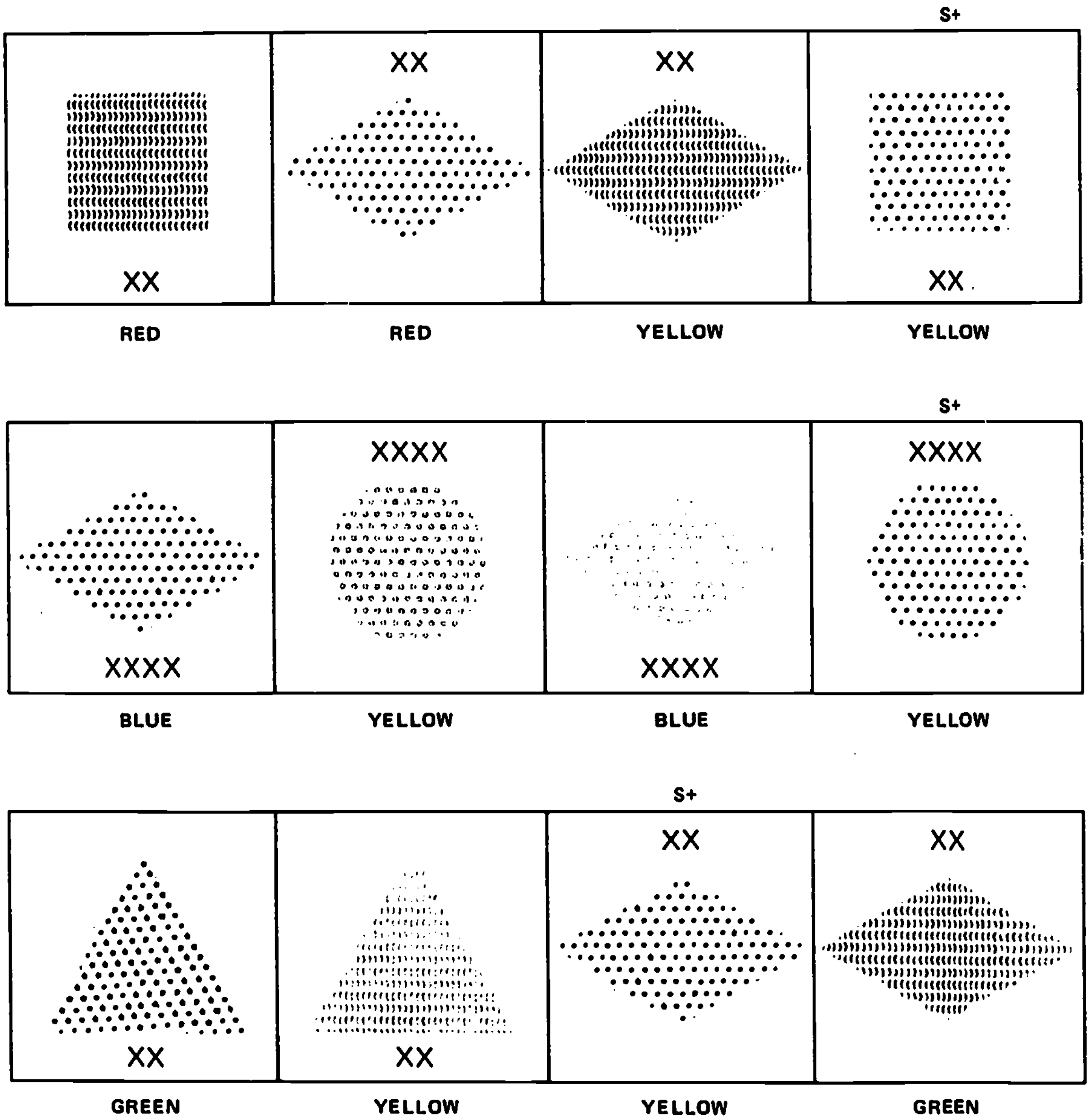

EXHIBIT 3 SAMPLE TRAINING ITEMS FOR SHIFT, CONCEPT 2

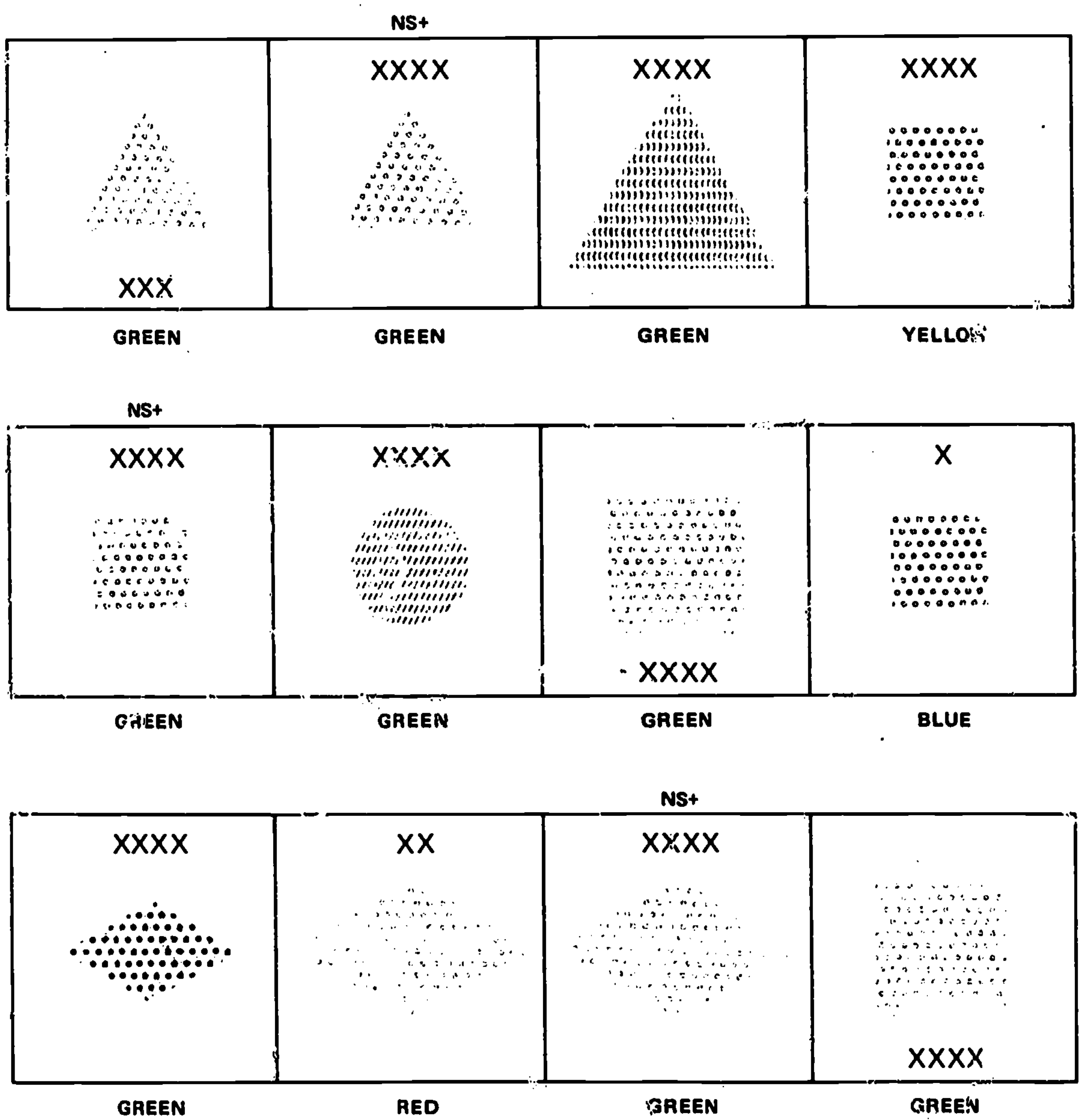

EXHIBIT 4 SAMPLE TRAINING ITEMS FOR NO-SHIFT, CONCEPT 3

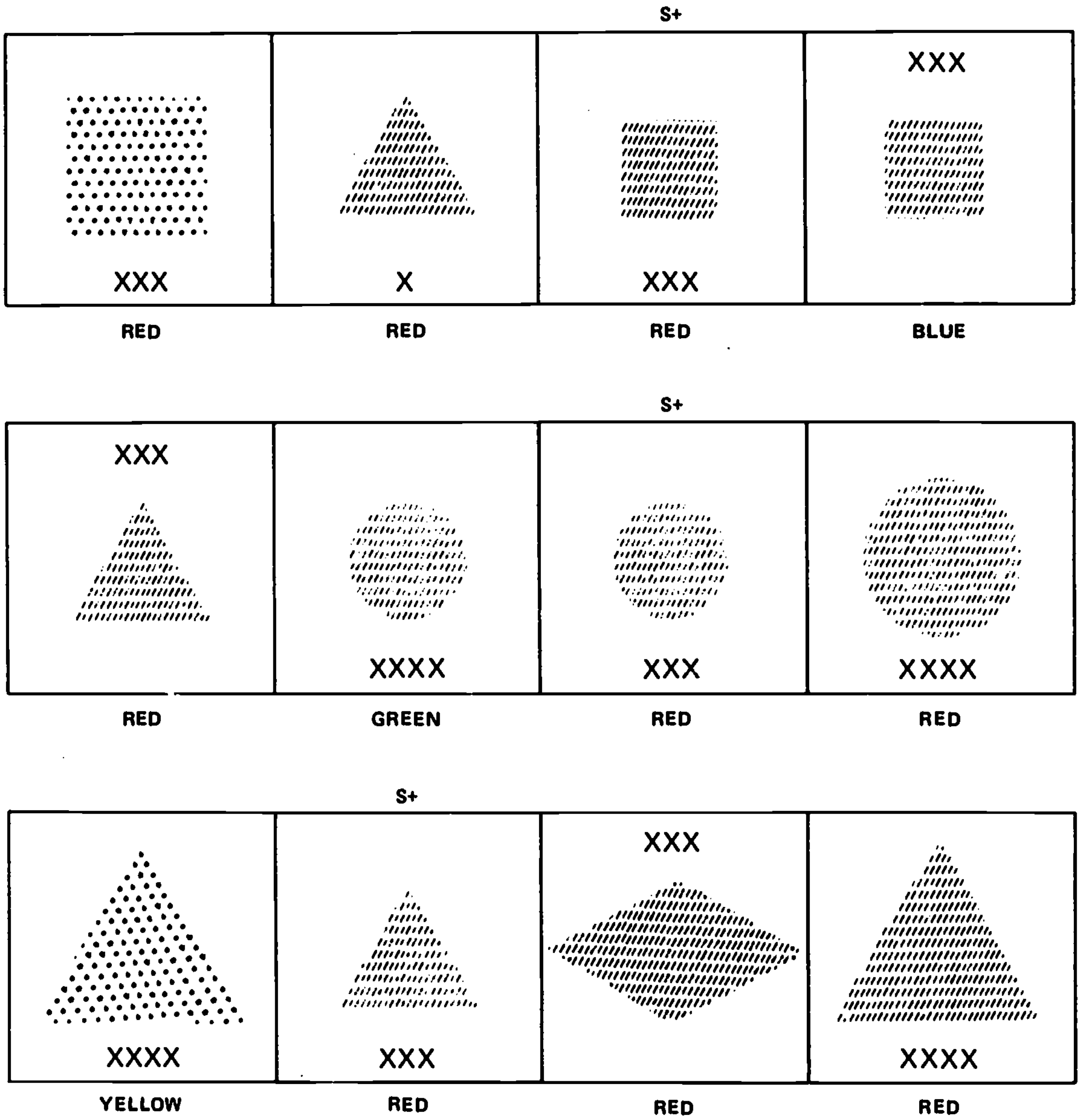

EXHIBIT 5 SAMPLE TRAINING ITEMS FOR SHIFT, CONCEPT 3

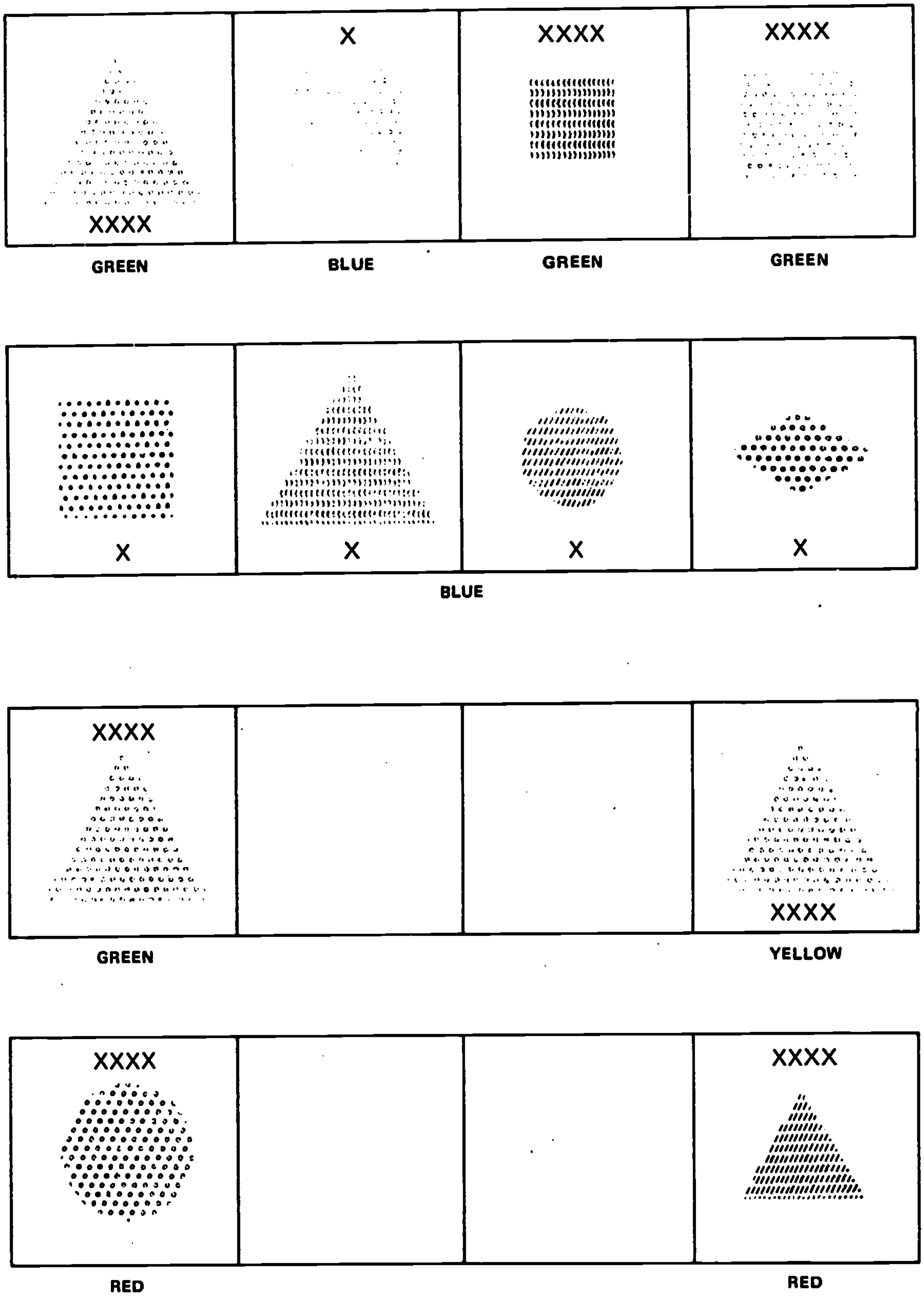

EXHIBIT 6 SAMPLE TEST ITEMS (Two and Four Choice)

NO.________

NAME____

SCHOOL______________________________________

GRADE______________CLASS____________________

AGE___________BOY GIRL

SAMPLE COVER SHEET FOR CONCEPT TRAINING RESPONSE BOOKLET

SAMPLE RESPONSE FORM FOR A SINGLE TRIAL IN CONCEPT TRAINING

EXHIBIT 7 SAMPLE COVER SHEET AND RESPONSE FORM FOR TRAINING

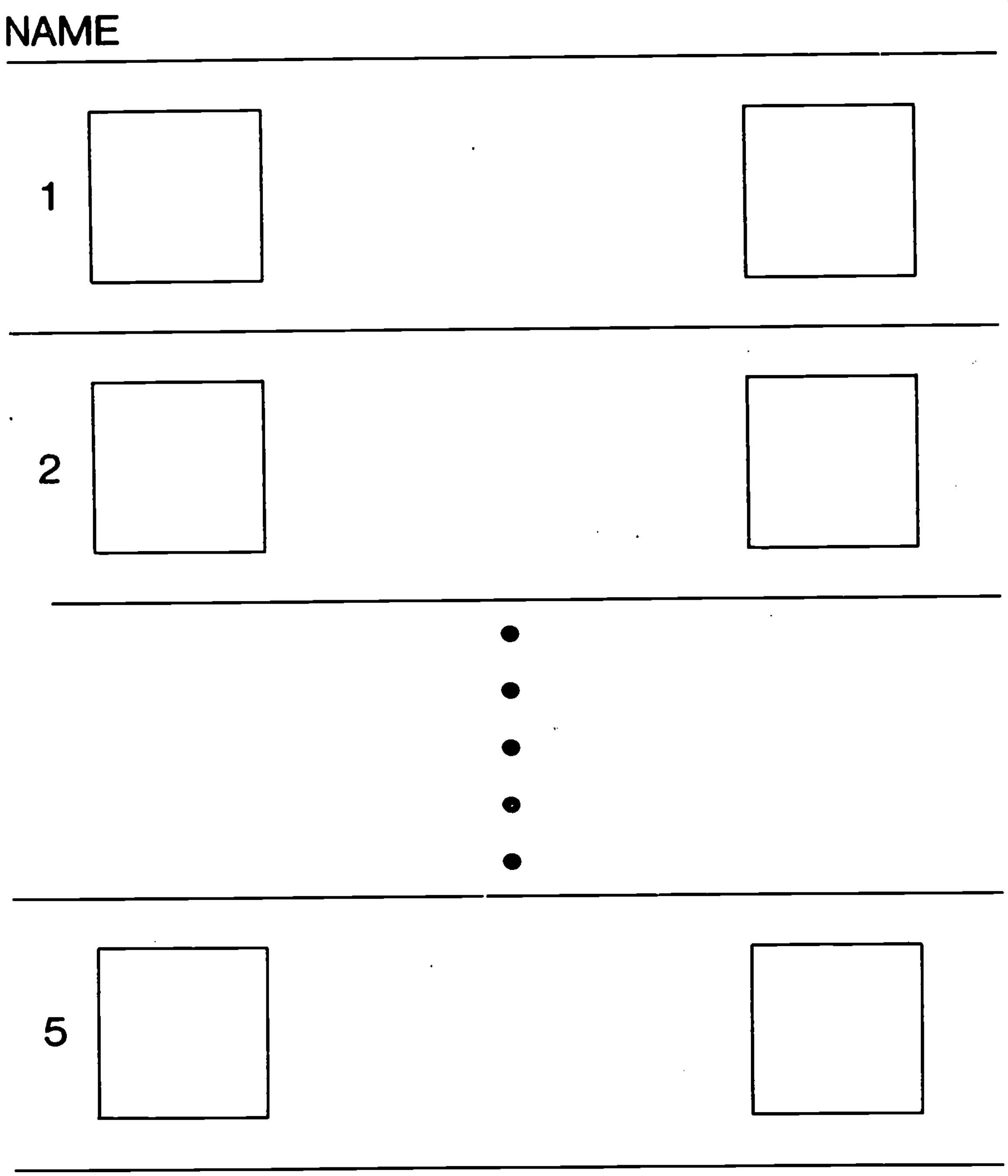

EXHIBIT 8 SAMPLE TEST RESPONSE FORM FOR TWO CHOICE TEST

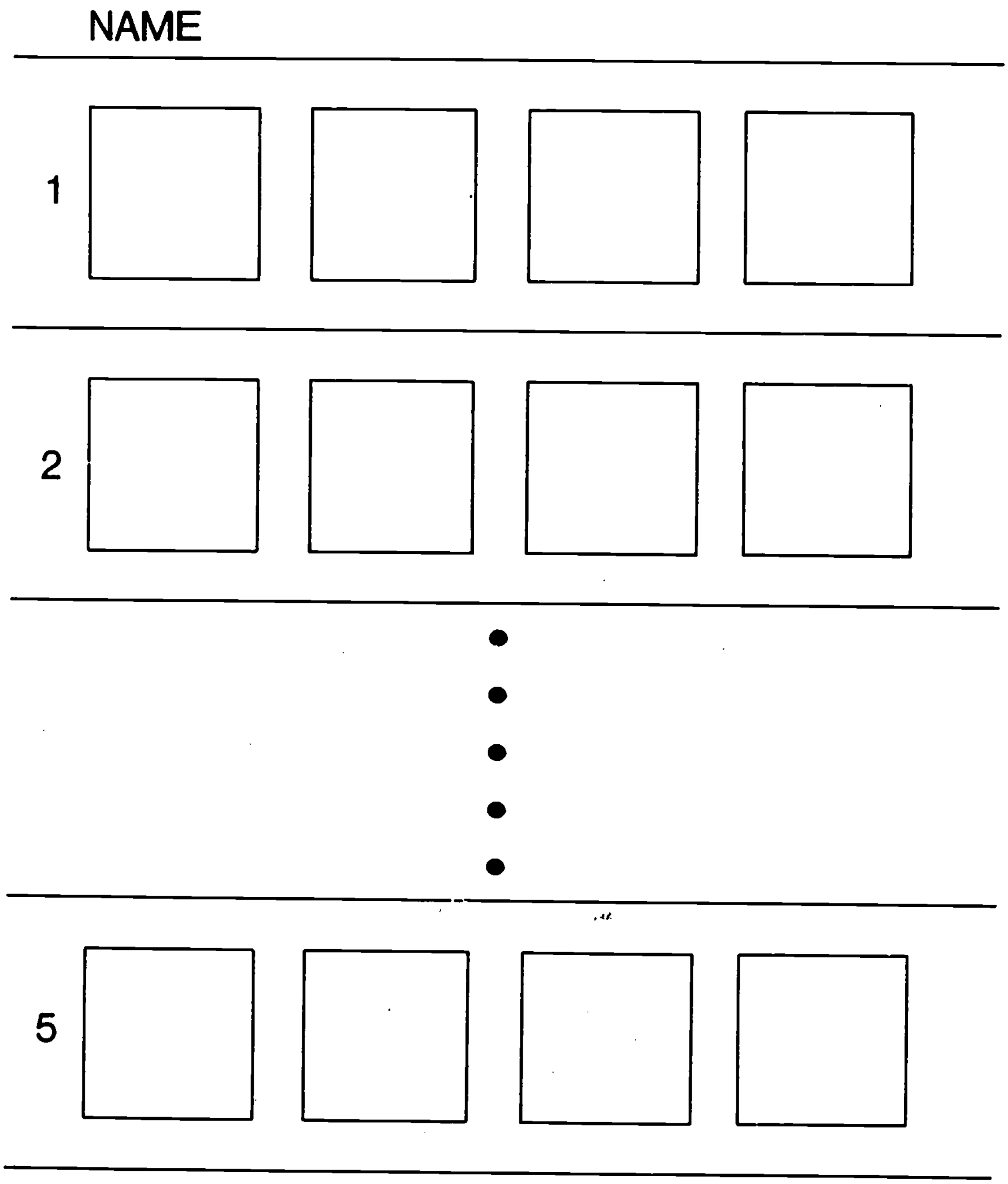

EXHIBIT 9 SAMPLE TEST RESPONSE FORM FOR FOUR CHOICE TEST

Appendix B

COMPUTER PROGRAM AND RESULTS OF ITEM ANALYSES AND MASTERY DECISION RULES

Appendix B

COMPUTER PROGRAM AND RESULTS OF ITEM ANALYSES AND MASTERY DECISION RULES

The contents of this appendix present the computer program used to compute the item analyses and the corresponding mastery decision rules for each of the 24 test groupings included in the experiment. These groupings were by test length (five or ten item) by item for (two or four choice) by concept (1, 2, or 3) for each of the two training paradigms.

The output for each of these test group item analyses includes item response protocol, item difficulty (percent pass), test mean, standard deviation, KR 20 reliability estimate, and estimated single item reliability.

Immediately following the item analysis output is listed a matrix of decision rules for the analyzed test, arranged in terms of mastery likelihood (ERR WT) and item type (ALPHA TO BETA RATIO). The encased values--a percent value and a corresponding "items to pass" value--correspond to the rules applied to the above test scores.

It should be noted that the cases and test scores are listed in pairs within training levels, such that the first two cases were trained under level 1, the next pair under level 2, and the third pair under level 3. This pattern is repeated for test groups based on 12 cases. Also, when the test reliability estimate approached values of zero or unity test cut rules were not calculated, and an $M/\overline{M}$ split of 50 percent was assumed.

```
              PROGRAM SCORE (INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
000003        DIMENSION PROB(10),X(30,30),S(30),P(30),N(10),FMT(12),
             1ID(30),KX(30,30),KP(30)
000003        DIMENSION WAB(9),WER(9), ALPHA(9),BETA(9),CJT(9,5),NCUT(9,5)
        C
        C  READ IN ALPHA/BETA AND ERROR WEIGHTS
        C
000003        READ(5,105) WAB
000011        READ(5,105) WER
000017    105 FORMAT (10F4.0)
        C
        C READ IN PROBLEM INFORMATION
        C
000017  1     READ(5,100) PROB,NC,NI
000031        IF(PROB(1).EQ.6HFINISH) GO TO 86
000033        WRITE(6,500) PROB,NC,NI
000045        READ(5,101) FMT
        C
        C   READ IN DATA VALUES FOR THIS PROBLEM. . INITIALIZE PARAMETERS
        C
000053        READ(5,FMT) (ID(I), (X(I,J),J=1,NI),I=1,NC)
000075        XN=NC
000077        SUM=0.
000100        SSQ=0.
000101  2     DO 10 J=1,NC
000103        S(J)=0.
        C
        C  CALCULATE TEST SCORES, MEAN AND VARIANCE
        C
000104        DO 11 I=1,NI
000106        KX(J,I) = X(J,I)
000114  11    S(J) = S(J) + X(J,I)
000121        SSQ=SSQ+S(J)**2
000123  10    SUM=SUM+S(J)
000127        XMEAN=SUM/XN
000131        VAR=((XN*SSQ)-SUM**2)/(XN*(XN-1.))
        C
        C  CALCULATE ITEM DIFFICULTIES (P AND Q).
        C
000137        SPQ=0.
000140        XI=NI
000141        DO 12 I=1,NI
000143        P(I)=0.
000144        DO 13 J=1,NC
000146  13    P(I)=P(I)+X(J,I)
000155        P(I)=P(I)/XN
000160        KP(I) = (P(I)*100.) +.5
000164        Q=1.0-P(I)
000166  12    SPQ=SPQ+Q*P(I)
        C
        C  CALCULATE TEST RELIABILITY (KR20).
        C
000173        SD = SQRT(VAR)
000175        RELIB = (XI/(XI-1.))*(1.-SPQ/VAR)
000203        SE = SD * SQRT (1.-RELIB)
        C
        C  CALCULATE ESTIMATED ITEM R (SH EST).
```

```
C
      WT=1./XI
      AVER=(WT*RELIB)/(1.+((WT-1.)*RELIB))
C
C  PRINT OUTPUT AND CHECK FOR NEXT PROBLEM.
C
      DO 14 K=1,10
14    N(K)=K
      WRITE(6,501) N
      DO 15 I=1,NC
      IF(NI.EQ.10) GO TO 16
      WRITE(6,502) ID(I),(KX(I,J),J=1,NI),S(I)
      GO TO 15
16    WRITE(6,503) ID(I),(KX(I,J),J=1,NI),S(I)
15    CONTINUE
      IF(NI.EQ.5) GO TO 20
      WRITE(6,504) (KP(I),I=1,NI),XMEAN,SD,RELIB,SE,AVER
      GO TO 150
 20   WRITE(6,505) (KP(I),I=1,NI),XMEAN,SD,RELIB,SE,AVER
  150 IF(RELIB.GE.1..OR.RELIB.LE.0.) GO TO 85
C
C CALCULATE CUT RULES
C
      VAL = 1. - SQRT(AVER)
      DO 75  I=1,9
      ALPHA(I) = VAL / (WAB(I)+1.)
      BETA(I) = VAL - ALPHA(I)
      CA = 1.- ALPHA(I)
      CB = 1. - BETA(I)
      DO 76  J = 1, 5
      CUT(I,J) = (ALOG(BETA(I)/CA)+(1./XN)*ALOG(WER(J)))/
     1  ALOG((ALPHA(I)*BETA(I))/(CA*CB))
      NCUT(I,J) = (CUT(I,J) * XI) + .5
   76 CONTINUE
   75 CONTINUE
C
C  PRINT  CUT-RULE OUTPUT
C
      WRITE(6,600) WAB
      DO 77  J = 1, 5
   77 WRITE(6,601) WER(J),(CUT(I,J),I=1,9),(NCUT(I,J),I=1,9)
      WRITE(6,602) VAL,(ALPHA(I),I=1,9),(BETA(I),I=1,9)
  602 FORMAT (*0ERROR       ALPHA/BETA VALUES*//F5.3,* ALPHA*
     1 F5.3,8F8.3/6X,* BETA* F5.3,8F8.3)
  600 FORMAT (*1 TEST CUT RULES*/20X,*ALPHA TO BETA RATIO* /
     1* ERR WT   * 9F8.3 /)
  601 FORMAT (*0*F6.2,2X,9F8.3/8X,9I8)
      GO TO 1
   85 WRITE(6,850)
  850 FORMAT(*0   CUT RULES NOT COMPUTED FOR THIS TEST.*)
      GO TO 1
86    WRITE (6,860)
      STOP
100   FORMAT(10A6,2I5)
101   FORMAT(12A5)
  500 FORMAT (*1             PROBLEM IDENTIFICATION...*10A6//
     1* NUMBER OF CASES . . . *I5/* NUMBER OF ITEMS . . . *I5/)
501   FORMAT(//* SUBJECT*10X,*I T E M *
```

```
                 114X,*TOTAL*/*  I.D.  *10I3,*  SCORE*/)
000536   502     FORMAT(2X,A6,5I3,15X,F6.1/)
000536   503     FORMAT(2X,A6,10I3,F6.1/)
000536   504     FORMAT(  *0PERCENT*/*  PASS  *10I3///
                 1* TEST MEAN . . . . . . . . * F6.2//
                 2* STANDARD DEVIATION  . . * F7.3//
                 3* RELIABILITY (KR20)  . . *F8.4//
                 4* STANDARD ERROR  . . . . * F8.4//
                 4* AVERAGE ITEM R  . . . . * F8.4/)
000536   505     FORMAT(  *0PERCENT*/*  PASS  * 5I3///
                 1* TEST MEAN . . . . . . . . * F6.2//
                 2* STANDARD DEVIATION  . . * F7.3//
                 3* RELIABILITY (KR20)  . . *F8.4//
                 4* STANDARD ERROR  . . . . * F8.4//
                 4* AVERAGE ITEM R  . . . . * F8.4/)
```

PROBLEM IDENTIFICATION...NO-SHIFT 2-OPT 5-ITEM CONCEPT 1 TEST (1.11.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
323	0	0	0	1	0						1.0
324	1	1	0	1	1						4.0
133	1	1	1	1	1						5.0
134	1	1	1	1	1						5.0
203	0	1	1	1	1						4.0
204	0	1	0	1	0						2.0
PERCENT PASS	50	83	50	100	67						

TEST MEAN 3.50

STANDARD DEVIATION . . 1.643

RELIABILITY (KR20) . . .8513

STANDARD ERROR 6336

AVERAGE ITEM R 5339

TEST CUT RULES

ERR.WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.128	.158	.190	.223	.294	.377	.427	.484	.553
	1	1	1	1	1	2	2	2	3
10.00	.208	.248	.286	.323	.397	.477	.523	.574	.633
	1	1	1	2	2	2	3	3	3
1.00	.287	.337	.382	.423	.500	.577	.619	.663	.713
	1	2	2	2	2	3	3	3	4
.10	.367	.426	.478	.523	.603	.677	.714	.752	.792
	2	2	2	3	3	3	4	4	4
.01	.447	.516	.574	.623	.706	.777	.810	.842	.872
	2	3	3	3	4	4	4	4	4

ERROR ALPHA/BETA VALUES

.269	ALPHA	.024	.045	.067	.090	.135	.180	.203	.224	.245
	BETA	.245	.224	.202	.180	.135	.090	.067	.045	.024

PROBLEM IDENTIFICATION...NO-SHIFT 2-OPT 5-ITEM CONCEPT 2 TEST (1.11.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
323	0	0	1	1	0						2.0
324	1	0	1	1	1						4.0
133	1	1	1	1	1						5.0
134	1	1	1	1	1						5.0
203	1	1	1	1	1						5.0
204	0	1	1	1	1						4.0

(ITEM)

PERCENT PASS 67 67100100 83

TEST MEAN 4.17

STANDARD DEVIATION . . 1.169

RELIABILITY (KR20) . . .7165

STANDARD ERROR 6225

AVERAGE ITEM R 3357

TEST CUT RULES

ALPHA TO BETA RATIO

ERR WT	10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.042 0	.065 0	.094 0	.128 1	.210 1	.315 2	.381 2	.456 2	.544 3
10.00	.145 1	.185 1	.226 1	.267 1	.355 2	.454 2	.512 3	.575 3	.647 3
1.00	.249 1	.305 2	.358 2	.407 2	.500 3	.593 3	.644 3	.695 3	.751 4
.10	.353 2	.425 2	.489 2	.546 3	.645 3	.733 4	.775 4	.815 4	.855 4
.01	.456 2	.544 3	.621 3	.685 3	.790 4	.872 4	.906 5	.935 5	.958 5

ERROR ALPHA/BETA VALUES

ERROR										
.421	ALPHA	.038	.073	.105	.140	.210	.280	.316	.350	.382
	BETA	.382	.350	.315	.280	.210	.140	.104	.070	.038

PROBLEM IDENTIFICATION...NO-SHIFT 2-OPT 5-ITEM CONCEPT 3 TEST (1.11.)

NUMBER OF CASES . . . 12
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
323	1	1	0	1	1						4.0
324	1	1	1	0	0						3.0
133	1	1	1	1	1						5.0
134	1	1	1	1	1						5.0
203	1	1	1	1	1						5.0
204	1	1	0	0	1						3.0
327	0	1	1	1	1						4.0
328	0	1	1	0	0						2.0
137	0	1	0	1	1						3.0
138	1	1	1	1	1						5.0
207	0	1	1	0	1						3.0
208	1	0	1	0	0						2.0
PERCENT PASS	67	92	75	58	75						

TEST MEAN 3.67

STANDARD DEVIATION . . 1.155

RELIABILITY (KR20) . . .3906

STANDARD ERROR 9014

AVERAGE ITEM R 1136

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.026 0	.049 0	.082 0	.123 1	.227 1	.363 2	.448 2	.541 3	.645 3
10.00	.108 1	.152 1	.200 1	.251 1	.363 2	.492 2	$\boxed{.566}$ 3	.644 3	.727 4
1.00	.191 1	.254 1	.318 2	.380 2	.506 2	.620 3	$\boxed{.683}$ 3	.746 4	.809 4
.10	.273 1	.356 2	.436 2	.508 3	.637 3	.749 4	$\boxed{.801}$ 4	.848 4	.892 4
.01	.355 2	.459 2	.554 3	.637 3	.773 4	.877 4	.919 5	.951 5	.974 5

ERROR ALPHA/BETA VALUES

.663 ALPHA	.060	.110	.166	.221	.331	.442	.498	.552	.603
BETA	.603	.552	.497	.442	.331	.221	.164	.110	.060

PROBLEM IDENTIFICATION...NO-SHIFT 4-UPT 5-ITEM CONCEPT I TEST (1.21.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
103	0	0	1	0	0						1.0
104	1	1	0	1	0						3.0
213	0	0	1	1	0						2.0
214	0	0	0	0	0						0.0
153	0	0	0	0	0						0.0
154	0	0	0	0	0						0.0
PERCENT PASS	17	17	33	33	0						

TEST MEAN 1.00

STANDARD DEVIATION . . 1.265

RELIABILITY (KR20) . . .6859

STANDARD ERROR 7091

AVERAGE ITEM R 3038

TEST CUT RULES

ALPHA TO BETA RATIO

ERR WT	10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.025 0	.046 0	.074 0	.107 1	.191 1	.300 2	.369 2	.448 2	.546 3
10.00	.133 1	.172 1	.213 1	.255 1	.345 2	.449 2	.509 3	.574 3	.649 3
1.00	.242 1	.299 1	.353 2	.401 2	.500 3	.597 3	.643 3	.701 4	.758 4
.10	.351 2	.426 2	.493 2	.552 3	.655 3	.745 4	.788 4	.828 4	.867 4
.01	.460 2	.552 3	.632 3	.700 3	.809 4	.893 4	.927 5	.954 5	.975 5

ERROR ALPHA/BETA VALUES

ERROR										
.449	ALPHA	.041	.075	.112	.150	.224	.299	.337	.374	.408
	BETA	.408	.374	.337	.299	.224	.150	.111	.075	.041

PROBLEM IDENTIFICATION...NO-SHIFT 4-OPT 5-ITEM CONCEPT 2 TEST (1.21.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
103	1	1	0	1	0						3.0
104	1	1	1	0	0						3.0
213	0	1	0	1	0						2.0
214	0	0	0	1	0						1.0
153	0	1	0	0	1						2.0
154	1	1	1	1	1						5.0
PERCENT PASS	50	83	33	67	33						

TEST MEAN 2.67

STANDARD DEVIATION . . 1.366

RELIABILITY (KR20) . . .5432

STANDARD ERROR 9235

AVERAGE ITEM R 1921

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	-.051 0	-.045 0	-.026 0	.004 0	.092 0	.221 1	.307 2	.404 2	.518 3
10.00	.082 0	.116 1	.155 1	.197 1	.296 1	.415 2	.487 2	.564 3	.652 3
1.00	.215 1	.276 1	.335 2	.391 2	.500 3	.609 3	.666 3	.724 4	.785 4
.10	.348 2	.436 2	.515 3	.585 3	.704 4	.803 4	.846 4	.884 4	.918 5
.01	.482 2	.596 3	.695 3	.779 4	.908 5	.996 5	1.026 5	1.045 5	1.051 5

ERROR ALPHA/BETA VALUES

ERROR										
.562	ALPHA	.051	.094	.140	.187	.281	.374	.422	.468	.511
	BETA	.511	.468	.421	.374	.281	.187	.139	.094	.051

PROBLEM IDENTIFICATION...NO-SHIFT 4-OPT 5-ITEM CONCEPT 3 TEST (1.21.)

NUMBER OF CASES . . . 12
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
103	0	0	0	0	0						0.0
104	0	0	0	0	0						0.0
213	0	0	0	0	0						0.0
214	0	0	1	0	0						1.0
153	0	0	0	0	0						0.0
154	1	1	1	1	1						5.0
108	0	0	0	0	1						1.0
107	1	0	0	0	0						1.0
217	1	0	0	0	0						1.0
218	0	0	0	0	0						0.0
157	1	1	0	1	0						3.0
158	1	1	1	1	1						5.0

PERCENT
PASS 42 25 25 25 25

TEST MEAN 1.42

STANDARD DEVIATION . . 1.881

RELIABILITY (KR20) . . .8991

STANDARD ERROR 5974

AVERAGE ITEM R 6407

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.237 1	.276 1	.313 2	.346 2	.413 2	.484 2	.524 3	.570 3	.623 3
10.00	.272 1	.315 2	.354 2	.389 2	.456 2	.526 3	.565 3	.608 3	.658 3
1.00	.307 2	.353 2	.395 2	.431 2	.500 3	.569 3	.606 3	.647 3	.693 3
.10	.342 2	.392 2	.436 2	.474 2	.544 3	.611 3	.647 3	.685 3	.728 4
.01	.377 2	.430 2	.477 2	.516 3	.587 3	.654 3	.688 3	.724 4	.763 4

ERROR ALPHA/BETA VALUES

ERROR										
.200	ALPHA	.018	.033	.050	.067	.100	.133	.150	.166	.181
	BETA	.181	.166	.150	.133	.100	.067	.050	.033	.018

PROBLEM IDENTIFICATION...NO-SHIFT 2-OPT 10-ITEM CONCEPT 1 TEST (1.12.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
321	0	0	1	0	1	0	0	0	1	0	3.0
322	1	0	1	0	0	1	0	0	0	1	4.0
131	1	1	1	0	0	0	1	1	0	1	6.0
132	1	1	1	1	1	1	1	1	1	1	10.0
201	0	0	0	0	0	0	0	0	0	0	0.0
202	1	1	1	1	1	1	1	1	1	1	10.0
PERCENT PASS	67	50	83	33	50	50	50	50	50	67	

TEST MEAN 5.50

STANDARD DEVIATION . . 3.987

RELIABILITY (KR20) . . .9500

STANDARD ERROR 8917

AVERAGE ITEM R 6552

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.173 2	.205 2	.236 2	.266 3	.329 3	.401 4	.444 4	.493 5	.553 6
10.00	.241 2	.280 3	.316 3	.349 3	.415 4	.484 5	.525 5	.569 6	.622 6
1.00	.310 3	.356 4	.396 4	.433 4	.500 5	.567 6	.605 6	.644 6	.690 7
.10	.378 4	.431 4	.476 5	.516 5	.585 6	.651 7	.685 7	.720 7	.759 8
.01	.447 4	.507 5	.557 6	.599 6	.671 7	.734 7	.765 8	.795 8	.827 8

ERROR ALPHA/BETA VALUES

ERROR										
.191	ALPHA	.017	.032	.048	.064	.095	.127	.143	.159	.173
	BETA	.173	.159	.143	.127	.095	.064	.047	.032	.017

PROBLEM IDENTIFICATION...NO-SHIFT 2-OPT 10-ITEM CONCEPT 2 TEST (1.12.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
321	0	0	1	1	0	0	1	1	0	0	4.0
322	0	0	1	1	0	0	1	1	1	0	5.0
131	1	1	1	0	0	1	1	0	0	1	6.0
132	1	1	1	1	1	1	1	1	1	1	10.0
201	0	1	0	0	1	0	1	1	1	1	6.0
202	1	1	1	1	1	0	0	0	0	0	5.0
PERCENT PASS	50	67	83	67	50	33	83	67	50	50	

TEST MEAN 6.00

STANDARD DEVIATION . . 2.098

RELIABILITY (KR20) . . .5640

STANDARD ERROR 1.3851

AVERAGE ITEM R 1145

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	-.137 -0	-.154 -1	-.151 -1	-.132 -0	-.045 0	.108 1	.215 2	.338 3	.481 5
10.00	.027 0	.050 1	.083 1	.124 1	.228 2	.364 4	.449 4	.542 5	.645 6
1.00	.191 2	.254 3	.318 3	.380 4	.500 5	.620 6	.683 7	.746 7	.809 8
.10	.355 4	.458 5	.553 6	.636 6	.772 8	.876 9	.917 9	.950 9	.973 10
.01	.519 5	.662 7	.788 8	.892 9	1.045 10	1.132 11	1.152 12	1.154 12	1.137 11

ERROR ALPHA/BETA VALUES

.662	ALPHA	.060	.110	.165	.221	.331	.441	.497	.551	.601
	BETA	.601	.551	.496	.441	.331	.221	.164	.110	.060

PROBLEM IDENTIFICATION...NO-SHIFT 2-OPT 10-ITEM CONCEPT 3 TEST (1.12.)

NUMBER OF CASES . . . 12
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
321	0	0	0	0	1	1	1	1	0	1	5.0
322	1	1	0	0	1	1	1	0	1	0	5.0
131	0	1	0	1	1	0	1	1	1	1	7.0
132	1	1	1	1	1	1	0	0	1	1	8.0
201	0	1	1	1	1	1	1	1	1	1	9.0
202	1	1	1	1	1	1	1	1	1	0	9.0
325	1	1	0	1	1	1	0	1	1	0	7.0
326	0	0	1	1	0	1	0	0	1	0	4.0
135	1	1	0	1	0	1	0	1	1	0	6.0
136	1	0	0	0	0	1	0	0	0	0	2.0
205	1	0	0	1	1	0	1	1	1	0	6.0
206	1	1	1	1	1	1	1	1	1	1	10.0
PERCENT PASS	67	67	42	75	75	83	58	67	83	42	

TEST MEAN 6.58

STANDARD DEVIATION . . 2.275

RELIABILITY (KR20) . . .6712

STANDARD ERROR 1.3043

AVERAGE ITEM R 1695

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.068 1	.100 1	.138 1	.181 2	.281 3	.404 4	.479 5	.560 6	.651 7
10.00	.139 1	.185 2	.234 2	.284 3	.390 4	.508 5	.575 6	.645 6	.721 7
1.00	.209 2	.270 3	.331 3	.388 4	.500 5	.612 6	.671 7	.730 7	.791 8
.10	.279 3	.355 4	.427 4	.492 5	.610 6	.716 7	.767 8	.815 8	.861 9
.01	.349 3	.440 4	.523 5	.596 6	.719 7	.819 8	.863 9	.900 9	.932 9

ERROR ALPHA/BETA VALUES

.588	ALPHA	.053	.098	.147	.196	.294	.392	.442	.490	.535
	BETA	.535	.490	.441	.392	.294	.196	.146	.098	.053

PROBLEM IDENTIFICATION...NO-SHIFT 4-OPT 10-ITEM CONCEPT 1 TEST (1.22.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
101	0	0	0	0	0	0	0	0	0	0	0.0
102	1	0	0	0	1	1	0	0	0	1	4.0
211	0	0	0	0	0	1	0	0	1	0	2.0
212	0	1	0	1	1	1	1	1	1	1	8.0
151	0	0	0	0	0	0	0	0	0	0	0.0
152	1	0	0	1	1	1	0	1	0	0	5.0
PERCENT PASS	33	17	0	33	50	67	17	33	33	33	

(Column heading "I T E M" spans items 1–10.)

TEST MEAN 3.17

STANDARD DEVIATION . . 3.125

RELIABILITY (KR20) . . .8994

STANDARD ERROR 9913

AVERAGE ITEM R 4720

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.103 1	.133 1	.165 2	.198 2	.272 3	.361 4	.416 4	.478 5	.552 6
10.00	.190 2	.230 2	.270 3	.308 3	.386 4	.472 5	.521 5	.575 6	.638 6
1.00	.276 3	.327 3	.375 4	.418 4	.500 5	.582 6	.626 6	.673 7	.724 7
.10	.362 4	.425 4	.480 5	.528 5	.614 6	.692 7	.731 7	.770 8	.810 8
.01	.448 4	.522 5	.585 6	.639 6	.728 7	.802 8	.836 8	.867 9	.897 9

ERROR ALPHA/BETA VALUES

ERROR										
.313	ALPHA	.028	.052	.078	.104	.156	.209	.235	.261	.285
	BETA	.285	.261	.235	.209	.156	.104	.078	.052	.028

PROBLEM IDENTIFICATION...NO-SHIFT 4-OPT 10-ITEM CONCEPT 2 TEST (1.22.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
101	1	0	0	0	1	0	0	1	0	0	3.0
102	0	0	1	0	0	0	0	0	0	0	1.0
211	1	1	1	0	0	1	0	0	0	0	4.0
212	1	1	1	0	0	1	1	0	1	1	7.0
151	1	1	0	1	0	1	1	1	1	0	7.0
152	1	1	1	1	1	1	1	1	1	1	10.0
PERCENT PASS	83	67	67	33	33	67	50	50	50	33	

TEST MEAN 5.33

STANDARD DEVIATION . . 3.266

RELIABILITY (KR20) . . .8796

STANDARD ERROR 1.1331

AVERAGE ITEM R 4222

TEST CUT RULES

ERR WT	10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
	ALPHA TO BETA RATIO								
100.00	.083 1	.110 1	.142 1	.175 2	.252 3	.347 3	.405 4	.471 5	.550 5
10.00	.174 2	.215 2	.255 3	.295 3	.376 4	.466 5	.519 5	.576 6	.642 6
1.00	.266 3	.319 3	.369 4	.414 4	.500 5	.586 6	.632 6	.681 7	.734 7
.10	.358 4	.424 4	.483 5	.534 5	.624 6	.705 7	.746 7	.785 8	.826 8
.01	.450 5	.529 5	.596 6	.653 7	.749 7	.825 8	.859 9	.890 9	.917 9

ERROR	ALPHA/BETA VALUES								
.350 ALPHA	.032	.058	.088	.117	.175	.233	.263	.292	.318
BETA	.318	.292	.263	.233	.175	.117	.087	.058	.032

PROBLEM IDENTIFICATION...NO-SHIFT 4-OPT 10-ITEM CONCEPT 3 TEST (1.22.)

NUMBER OF CASES . . . 12
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
101	0	0	0	0	0	0	0	0	0	0	0.0
102	0	0	0	0	0	0	0	0	0	0	0.0
211	1	1	1	1	1	1	1	1	1	1	10.0
212	1	1	1	1	1	1	1	1	1	1	10.0
151	1	0	1	1	0	1	1	0	0	0	5.0
152	1	1	1	1	1	1	1	1	1	1	10.0
105	0	0	0	0	0	0	0	0	0	0	0.0
106	0	0	1	0	0	0	0	0	0	0	1.0
215	0	0	0	0	0	0	1	0	0	0	1.0
226	0	0	0	0	0	0	0	1	0	0	1.0
155	1	1	1	1	1	1	1	1	1	1	10.0
156	0	0	0	0	0	1	0	0	0	0	1.0
PERCENT PASS	42	33	50	42	33	50	50	42	33	33	

TEST MEAN 4.08

STANDARD DEVIATION . . 4.562

RELIABILITY (KR20) . . .9847

STANDARD ERROR 5647

AVERAGE ITEM R 8653

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.306 3	.339 3	.368 4	.394 4	.442 4	.493 5	.522 5	.555 6	.595 6
10.00	.330 3	.365 4	.396 4	.422 4	.471 5	.521 5	.550 5	.581 6	.620 6
1.00	.355 4	.392 4	.423 4	.451 5	.500 5	.549 5	.577 6	.608 6	.645 6
.10	.380 4	.419 4	.451 5	.479 5	.529 5	.578 6	.605 6	.635 6	.670 7
.01	.405 4	.445 4	.479 5	.507 5	.558 6	.606 6	.633 6	.661 7	.694 7

ERROR ALPHA/BETA VALUES

ERROR										
.070	ALPHA	.006	.012	.017	.023	.035	.047	.052	.058	.063
	BETA	.063	.058	.052	.047	.035	.023	.017	.012	.006

PROBLEM IDENTIFICATION...SHIFT 2-OPTION 5-ITEM CONCEPT 1 TEST (2.11.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
113	0	1	1	1	1						4.0
114	1	1	1	1	1						5.0
313	1	1	1	0	1						4.0
314	1	1	1	1	1						5.0
303	1	1	1	1	1						5.0
304	1	1	1	1	1						5.0
PERCENT PASS	83	100	100	83	100						

TEST MEAN 4.67

STANDARD DEVIATION . . .516

RELIABILITY (KR20) . . -.0521

STANDARD ERROR 5297

AVERAGE ITEM R . . . . -.0100

CUT RULES NOT COMPUTED FOR THIS TEST.

89

PROBLEM IDENTIFICATION...SHIFT 2-OPTION 5-ITEM CONCEPT 2 TEST (2.11.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
113	1	1	1	1	1						5.0
114	1	1	0	1	0						3.0
313	1	0	1	1	0						3.0
314	0	1	1	0	1						3.0
303	1	1	1	1	1						5.0
304	1	1	1	1	1						5.0
PERCENT PASS	83	83	83	83	67						

TEST MEAN 4.00

STANDARD DEVIATION . . 1.095

RELIABILITY (KR20) . . .4398

STANDARD ERROR 8199

AVERAGE ITEM R 1357

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	-.109 -0	-.116 -0	-.108 -0	-.085 0	.004 0	.149 1	.248 1	.362 2	.495 2
10.00	.045 0	.072 0	.108 1	.149 1	.252 1	.383 2	.463 2	.551 3	.548 3
1.00	.198 1	.261 1	.323 2	.383 2	.500 3	.617 3	.678 3	.739 4	.802 4
.10	.352 2	.449 2	.539 3	.617 3	.744 4	.851 4	.893 4	.928 5	.955 5
.01	.505 3	.638 3	.755 4	.851 4	.996 5	1.085 5	1.108 6	1.116 6	1.109 6

ERROR ALPHA/BETA VALUES

.632	ALPHA	.057	.105	.158	.211	.316	.421	.475	.526	.574
	BETA	.574	.526	.474	.421	.316	.211	.157	.105	.057

PROBLEM IDENTIFICATION...SHIFT 2-OPTION 5-ITEM CONCEPT 3 TEST (2.11.)

NUMBER OF CASES . . . 12
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
113	0	0	0	0	0						0.0
114	0	0	1	0	0						1.0
313	0	0	1	1	0						2.0
314	1	0	1	1	0						3.0
303	1	1	1	1	1						5.0
304	1	0	1	0	1						3.0
127	0	0	0	0	0						0.0
128	0	1	0	1	1						3.0
317	0	1	0	0	0						1.0
318	0	1	1	0	0						2.0
307	1	0	1	1	1						4.0
308	0	0	0	0	1						1.0
PERCENT PASS	33	33	58	42	42						

TEST MEAN 2.08

STANDARD DEVIATION . . 1.564

RELIABILITY (KR20) . . .6505

STANDARD ERROR 9248

AVERAGE ITEM R 2712

TEST CUT RULES

ALPHA TO BETA RATIO

ERR WT	10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.120	.158	.199	.241	.334	.441	.504	.573	.650
	1	1	1	1	2	2	3	3	3
10.00	.178	.225	.274	.321	.417	.520	.579	.640	.708
	1	1	1	2	2	3	3	3	4
1.00	.235	.293	.348	.400	.500	.600	.653	.707	.765
	1	1	2	2	3	3	3	4	4
.10	.292	.360	.423	.480	.583	.679	.727	.775	.822
	1	2	2	2	3	3	4	4	4
.01	.350	.427	.497	.559	.666	.759	.802	.842	.880
	2	2	2	3	3	4	4	4	4

ERROR ALPHA/BETA VALUES

.479	ALPHA	.044	.080	.120	.160	.240	.319	.360	.399	.436
	BETA	.436	.399	.359	.319	.240	.160	.119	.080	.044

PROBLEM IDENTIFICATION...SHIFT 4-OPTION 5-ITEM CONCEPT 1 TEST (2.21.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
123	1	1	1	1	1						5.0
124	0	0	1	0	1						2.0
143	0	0	0	0	0						0.0
144	1	1	1	1	1						5.0
223	1	0	0	0	0						1.0
224	1	1	1	1	1						5.0
PERCENT PASS	67	50	67	50	67						

TEST MEAN 3.00

STANDARD DEVIATION . . 2.280

RELIABILITY (KR20) . . .9696

STANDARD ERROR 3979

AVERAGE ITEM R 8643

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.256 1	.285 1	.312 2	.337 2	.384 2	.436 2	.467 2	.502 3	.546 3
10.00	.305 2	.339 2	.368 2	.393 2	.442 2	.493 2	.522 3	.555 3	.595 3
1.00	.355 2	.392 2	.423 2	.459 2	.500 3	.550 3	.577 3	.608 3	.645 3
.10	.405 2	.445 2	.479 2	.507 3	.558 3	.607 3	.633 3	.661 3	.695 3
.01	.454 2	.498 2	.534 3	.564 3	.616 3	.663 3	.688 3	.715 4	.744 4

ERROR ALPHA/BETA VALUES

ERROR										
.070	ALPHA	.006	.012	.018	.023	.035	.047	.053	.059	.064
	BETA	.064	.059	.053	.047	.035	.023	.017	.012	.006

PROBLEM IDENTIFICATION...SHIFT 4-OPTION 5-ITEM CONCEPT 2 TEST (2.21.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
123	1	1	1	1	1						5.0
124	0	1	1	1	1						4.0
143	1	0	0	0	1						2.0
144	1	1	1	1	1						5.0
223	1	1	0	1	0						3.0
224	1	1	1	1	1						5.0
PERCENT PASS	83	83	67	83	83						

TEST MEAN 4.00

STANDARD DEVIATION . . 1.265

RELIABILITY (KR20) . . .6424

STANDARD ERROR 7565

AVERAGE ITEM R 2643

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.001 0	.019 0	.044 0	.077 0	.162 1	.278 1	.352 2	.436 2	.535 3
10.00	.117 1	.155 1	.196 1	.238 1	.331 2	.439 2	.503 3	.572 3	.651 3
1.00	.233 1	.291 1	.347 2	.399 2	.500 3	.601 3	.654 3	.709 4	.767 4
.10	.349 2	.428 2	.499 2	.561 3	.669 3	.762 4	.805 4	.845 4	.883 4
.01	.465 2	.564 3	.650 3	.722 4	.834 4	.923 5	.956 5	.981 5	.999 5

ERROR ALPHA/BETA VALUES

ERROR										
.486	ALPHA	.044	.081	.121	.162	.243	.324	.365	.405	.442
	BETA	.442	.405	.364	.324	.243	.162	.121	.081	.044

PROBLEM IDENTIFICATION...SHIFT 4-OPTION 5-ITEM CONCEPT 3 TEST (2.21.)

NUMBER OF CASES . . . 12
NUMBER OF ITEMS . . . 5

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
123	1	0	1	1	1						4.0
124	0	1	1	1	1						4.0
143	0	0	0	0	0						0.0
144	1	1	1	1	1						5.0
223	1	0	0	1	1						3.0
224	1	0	1	1	1						4.0
117	1	0	0	1	1						3.0
118	1	0	1	0	0						2.0
147	1	0	0	1	0						2.0
148	0	0	0	0	0						0.0
227	1	0	0	1	1						3.0
228	0	1	1	0	0						2.0
PERCENT PASS	67	25	50	67	58						

TEST MEAN 2.67

STANDARD DEVIATION . . 1.557

RELIABILITY (KR20) . . .6699

STANDARD ERROR 8945

AVERAGE ITEM R 2887

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.127 1	.166 1	.207 1	.249 1	.340 2	.445 2	.507 3	.574 3	.650 3
10.00	.183 1	.231 1	.279 1	.326 2	.420 2	.522 3	.578 3	.639 3	.705 4
1.00	.239 1	.296 1	.351 2	.402 2	.500 3	.598 3	.650 3	.704 4	.761 4
.10	.295 1	.361 2	.423 2	.478 2	.580 3	.674 3	.722 4	.769 4	.817 4
.01	.350 2	.426 2	.495 2	.555 3	.660 3	.751 4	.794 4	.834 4	.873 4

ERROR ALPHA/BETA VALUES

ERROR										
.463	ALPHA	.042	.077	.116	.154	.231	.308	.348	.386	.421
	BETA	.421	.386	.347	.308	.231	.154	.115	.077	.042

PROBLEM IDENTIFICATION...SHIFT 2-OPTION 10-ITEM CONCEPT 1 TEST (2.12.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
121	1	1	0	0	0	0	0	0	1	1	4.0
122	0	0	1	0	1	0	0	1	0	1	4.0
311	1	1	1	1	1	1	1	1	1	1	10.0
312	1	1	1	1	1	1	1	1	1	1	10.0
301	1	1	1	1	1	1	1	1	1	1	10.0
302	1	1	1	1	1	1	1	1	1	1	10.0
PERCENT PASS	83	83	83	67	83	67	67	83	83	100	

TEST MEAN 8.00

STANDARD DEVIATION . . 3.098

RELIABILITY (KR20) . . .9375

STANDARD ERROR 7746

AVERAGE ITEM R 6000

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.153 2	.184 2	.216 2	.247 2	.314 3	.391 4	.437 4	.490 5	.554 6
10.00	.226 2	.266 3	.303 3	.338 3	.407 4	.481 5	.524 5	.571 6	.627 6
1.00	.300 3	.347 3	.390 4	.428 4	.500 5	.572 6	.611 6	.653 7	.700 7
.10	.373 4	.429 4	.477 5	.519 5	.593 6	.662 7	.698 7	.734 7	.774 8
.01	.446 4	.510 5	.564 6	.609 6	.686 7	.753 8	.785 8	.816 8	.847 8

ERROR ALPHA/BETA VALUES

ERROR										
.225	ALPHA	.020	.038	.056	.075	.113	.150	.169	.188	.205
	BETA	.205	.188	.169	.150	.113	.075	.056	.038	.020

95

PROBLEM IDENTIFICATION...SHIFT 2-OPTION 10-ITEM CONCEPT 2 TEST (2.12.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
121	1	1	0	0	0	1	0	0	0	1	4.0
122	1	1	1	1	1	0	1	1	0	1	8.0
311	1	1	1	1	1	1	1	1	1	1	10.0
312	0	0	0	0	0	0	0	0	0	0	0.0
301	1	1	1	1	1	1	1	1	1	1	10.0
302	1	1	1	1	1	1	1	1	1	1	10.0
PERCENT PASS	83	83	67	67	67	67	67	67	50	83	

TEST MEAN 7.00

STANDARD DEVIATION . . 4.147

RELIABILITY (KR20) . . .9819

STANDARD ERROR 5578

AVERAGE ITEM R 8444

TEST CUT RULES

ALPHA TO BETA RATIO

ERR WT	10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.247 2	.277 3	.304 3	.329 3	.379 4	.433 4	.465 5	.501 5	.547 5
10.00	.298 3	.332 3	.362 4	.389 4	.439 4	.492 5	.523 5	.557 6	.599 6
1.00	.350 3	.388 4	.420 4	.448 4	.500 5	.552 6	.581 6	.612 6	.650 7
.10	.401 4	.443 4	.478 5	.508 5	.561 6	.611 6	.638 6	.668 7	.702 7
.01	.453 5	.499 5	.536 5	.567 6	.621 6	.671 7	.696 7	.723 7	.753 8

ERROR ALPHA/BETA VALUES

.081	ALPHA	.007	.014	.020	.027	.041	.054	.061	.068	.074
	BETA	.074	.068	.061	.054	.041	.027	.020	.014	.007

PROBLEM IDENTIFICATION...SHIFT 2-OPTION 10-ITEM CONCEPT 3 TEST (2.12.)

NUMBER OF CASES . . . 12
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
121	0	0	0	1	1	1	1	0	1	0	5.0
122	1	0	0	1	1	1	0	1	0	1	6.0
311	1	0	1	1	0	1	1	1	1	0	7.0
312	1	1	0	0	1	1	1	0	0	1	6.0
301	1	1	1	1	1	1	1	1	1	1	10.0
302	1	1	1	0	0	0	0	0	0	0	3.0
125	0	1	1	1	1	1	0	1	1	0	7.0
126	0	0	0	1	1	1	0	0	0	1	4.0
315	0	1	1	0	1	0	0	1	1	0	5.0
316	1	0	1	1	1	1	1	0	0	0	6.0
305	1	0	0	0	0	1	1	1	1	0	5.0
306	0	1	1	1	1	1	1	0	1	0	7.0
PERCENT PASS	58	50	58	67	75	83	58	50	58	33	

TEST MEAN 5.92

STANDARD DEVIATION . . 1.782

RELIABILITY (KR20) . . .3200

STANDARD ERROR 1.4627

AVERAGE ITEM R 0461

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	-.073 -0	-.060 -0	-.069 -0	-.042 0	.060 1	.228 2	.341 3	.468 5	.608 6
10.00	.043 0	.073 1	.113 1	.161 2	.280 3	.432 4	.523 5	.621 6	.725 7
1.00	.159 2	.226 2	.296 3	.365 4	.500 5	.635 6	.705 7	.774 8	.841 8
.10	.275 3	.379 4	.479 5	.568 6	.720 7	.839 8	.888 9	.927 9	.957 10
.01	.392 4	.532 5	.662 7	.772 8	.940 9	1.042 10	1.070 11	1.080 11	1.073 11

ERROR ALPHA/BETA VALUES

ERROR										
.785	ALPHA	.071	.131	.196	.262	.393	.523	.590	.654	.714
	BETA	.714	.654	.589	.523	.393	.262	.195	.131	.071

PROBLEM IDENTIFICATION...SHIFT 4-OPTION 10-ITEM CONCEPT 1 TEST (2.22.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
111	1	1	1	1	1	1	1	1	1	1	10.0
112	0	0	1	1	0	0	0	0	0	0	2.0
141	1	1	1	1	1	1	1	1	1	1	10.0
142	0	0	0	0	0	1	0	1	0	0	2.0
221	1	1	1	1	1	1	1	1	1	1	10.0
222	1	1	1	1	1	1	1	1	1	1	10.0
PERCENT PASS	67	67	83	83	67	83	67	83	67	67	

TEST MEAN 7.33

STANDARD DEVIATION . . 4.131

RELIABILITY (KR20) . . .9881

STANDARD ERROR 4500

AVERAGE ITEM R 8928

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.270	.299	.324	.348	.392	.441	.469	.502	.544
	3	3	3	3	4	4	5	5	5
10.00	.316	.348	.376	.400	.446	.494	.521	.552	.590
	3	3	4	4	4	5	5	6	6
1.00	.363	.398	.428	.453	.500	.547	.573	.602	.637
	4	4	4	5	5	5	6	6	6
.10	.410	.448	.480	.506	.554	.600	.625	.652	.684
	4	4	5	5	6	6	6	7	7
.01	.456	.498	.531	.559	.608	.652	.676	.701	.730
	5	5	5	6	6	7	7	7	7

ERROR ALPHA/BETA VALUES

ERROR										
.055	ALPHA	.005	.009	.014	.018	.028	.037	.041	.046	.050
	BETA	.050	.046	.041	.037	.028	.018	.014	.009	.005

PROBLEM IDENTIFICATION...SHIFT 4-OPTION 10-ITEM CONCEPT 2 TEST (2.22.)

NUMBER OF CASES . . . 6
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
	ITEM										
111	1	1	1	1	1	1	1	1	1	1	10.0
112	1	0	0	0	0	0	0	1	1	0	3.0
141	1	1	1	1	1	1	1	1	1	1	10.0
142	1	0	0	0	0	1	0	0	1	0	3.0
221	1	1	1	1	1	1	1	1	1	1	10.0
222	1	1	1	1	1	0	1	1	1	1	9.0
PERCENT PASS	100	67	67	67	67	67	67	83	100	67	

TEST MEAN 7.50

STANDARD DEVIATION . . 3.507

RELIABILITY (KR20) . . .9580

STANDARD ERROR 7184

AVERAGE ITEM R 6954

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.188 2	.219 2	.250 3	.280 3	.340 3	.408 4	.449 4	.496 5	.552 6
10.00	.253 3	.291 3	.326 3	.358 4	.420 4	.486 5	.525 5	.567 6	.617 6
1.00	.318 3	.362 4	.401 4	.436 4	.500 5	.564 6	.600 6	.638 6	.682 7
.10	.383 4	.433 4	.476 5	.514 5	.580 6	.642 6	.675 7	.709 7	.747 7
.01	.448 4	.504 5	.552 6	.592 6	.660 7	.720 7	.751 8	.781 8	.812 8

ERROR	ALPHA/BETA VALUES									
.166	ALPHA	.015	.028	.042	.055	.083	.111	.125	.138	.151
	BETA	.151	.138	.125	.111	.083	.055	.041	.028	.015

PROBLEM IDENTIFICATION...SHIFT 4-OPTION 10-ITEM CONCEPT 3 TEST (2.22.)

NUMBER OF CASES . . . 12
NUMBER OF ITEMS . . . 10

SUBJECT I.D.	ITEM 1	2	3	4	5	6	7	8	9	10	TOTAL SCORE
111	0	0	1	1	1	1	1	1	0	1	7.0
112	0	0	1	0	0	0	0	0	0	0	1.0
141	0	0	0	1	0	0	0	0	0	0	1.0
142	0	0	0	0	0	0	0	0	0	0	0.0
221	1	0	1	0	1	1	0	0	0	1	5.0
222	0	0	0	0	0	0	0	1	1	0	2.0
115	0	0	0	0	0	0	0	0	0	0	0.0
116	1	1	0	0	0	1	1	1	1	1	7.0
145	0	1	0	0	0	1	1	1	1	1	6.0
146	0	1	0	0	0	0	0	0	0	1	2.0
225	1	0	0	0	0	0	1	0	0	1	3.0
226	0	0	0	0	0	0	0	0	0	0	0.0
PERCENT PASS	25	25	25	17	17	33	33	33	25	50	

TEST MEAN 2.83

STANDARD DEVIATION . . 2.725

RELIABILITY (KR20) . . .8201

STANDARD ERROR 1.1557

AVERAGE ITEM R 3131

TEST CUT RULES

ERR WT	ALPHA TO BETA RATIO 10.000	5.000	3.000	2.000	1.000	.500	.330	.200	.100
100.00	.137 1	.176 2	.217 2	.259 3	.348 3	.450 5	.510 5	.575 6	.649 6
10.00	.191 2	.238 2	.286 3	.332 3	.424 4	.523 5	.578 6	.637 6	.702 7
1.00	.244 2	.301 3	.354 4	.404 4	.500 5	.596 6	.647 6	.699 7	.756 8
.10	.298 3	.363 4	.423 4	.477 5	.576 6	.668 7	.715 7	.762 8	.809 8
.01	.351 4	.425 4	.492 5	.550 5	.652 7	.741 7	.784 8	.824 8	.863 9

ERROR ALPHA/BETA VALUES

.440	ALPHA	.040	.073	.110	.147	.220	.294	.331	.367	.400
	BETA	.400	.367	.330	.294	.220	.147	.109	.073	.040